

Remodeling our System of Assessments in New Mexico

Current Conditions
and Opportunities

WHITE PAPER
OCTOBER 2018



Prepared by Deborah Good, Future Focused Education
In support of a project funded by the Assessment for Learning Project

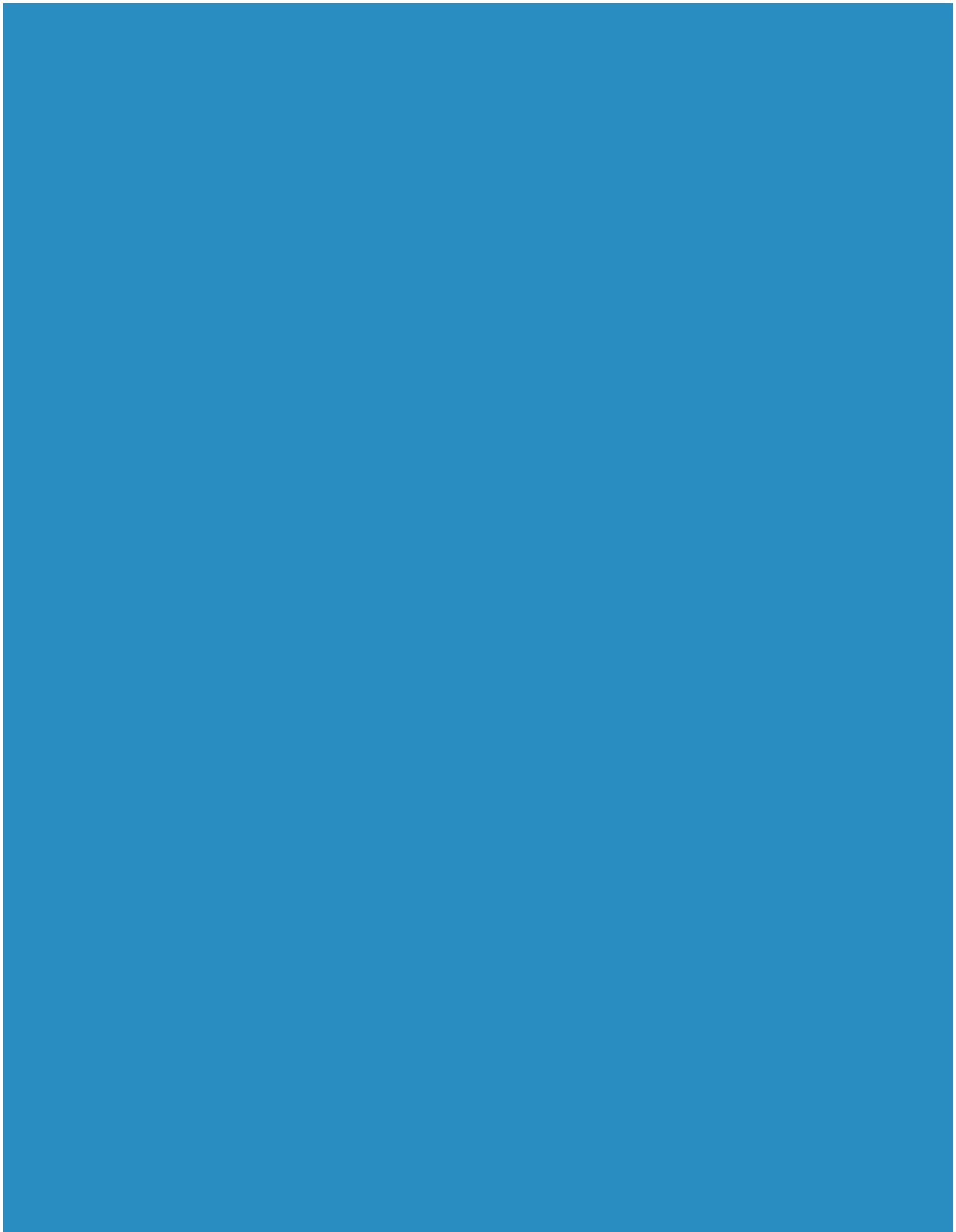


Table of Contents

Dear Reader	i
Introduction	i
Why Change Assessment?	3
What Does a New Vision Look Like?	5
What principles might guide a new assessment system in New Mexico?	5
What kinds of assessment are there?	5
What is a system of assessments?	8
What Are the Barriers to Change?	10
Arena 1: Federal Policy	10
Arena 2: State Policy	11
Arena 3: Culture of Accountability	12
Arena 4: Educator Capacity	13
Where Do We Start?	14
Foothold 1: Graduate Profile	14
Foothold 2: Grades and Subjects with Less Required Testing	15
Foothold 3: Career and Technical Education	16
What National Bright Spots Can We Learn From?	17
New Hampshire	17
Virginia	17
Hawaii	17
Conclusion	18
Appendix A: Proposed Theory of Change	19
Appendix B. Central New Mexico Graduate Profile	20
Acknowledgements	22

Dear Reader:

“We can standardize tests but we can’t standardize children. We must seek out ways to develop our students’ unique skills and perspectives, and give them authentic choice in how they will demonstrate their learning.”

Dr. Donna Navarrete,
Turnaround Specialist, Cuba
Independent School District

Over my years as a school leader and education advocate, I have witnessed repeated changes to assessment policy and practice in New Mexico. With each new iteration, we saw a huge public investment and many positive intentions. Debates erupted over the pros and cons of new systems. Meanwhile, young people’s needs and voices were rarely included, local wisdom was frequently overshadowed, and educators’ expertise was too often ignored.

Despite our well-intentioned efforts and endless clamoring across the education landscape, student learning has not improved in significant ways. National assessments have shown little change in our students’ performance over the last two decades.¹

We are entering another season of transition in New Mexico, and several factors make this an opportune time for a statewide conversation about student assessment:

- The new governor plans to end use of the controversial PARCC exam and stop “punitive testing-first evaluation.”²
- A School Support & Accountability Work Group convened by New Mexico’s Legislative Education Study Committee is recommending “an assessment system that supports assessment for student learning.”³
- A team of district leaders and Future Focused Education had a breakthrough on a recent trip to New Hampshire, where performance assessments are used in valid and reliable ways in the state’s accountability system. This answers much of the skepticism about moving away from standardized testing as we know it.
- The one-size-fits-all, single-state-test accountability system has not served New Mexico’s diverse population well over the last 20 years, and the federal Every Student Succeeds Act (ESSA) opens doors for flexibility in assessment that New Mexico has not yet utilized to build a comprehensive system of assessment.
- The recent *Yazzie/Martinez vs the State of New Mexico* decision sets the stage for transforming support for all students, especially students with disabilities and students of color.
- In the face of extreme teacher shortages, New Mexico must recommit to valuing teachers as professionals.

¹ See New Mexico scores on the national NAEP exam: <https://www.nationsreportcard.gov/>

² New Mexicans for Michelle. *A New Direction for New Mexico’s Public Schools* [Webpage]. Retrieved from <http://www.newmexicansformichelle.com/policy/a-new-direction-for-new-mexico-s-public-schools>

³ School Support and Accountability Work Group (2018, November 14). *Senate Memorial 145 – School Grades Work Group: Report to the Legislative Education Study Committee* [New Mexico Legislative Committee Report], p. 2.

This white paper was prepared in October 2018 to support stakeholders in imagining initial steps and a long-term vision for a new system of assessments in New Mexico. It does not prescribe a particular reform strategy but provides a foundation for further learning and decision-making. Since releasing an initial draft of the paper, our team has gathered stakeholder feedback, traveled with a group of district and state leaders to New Hampshire, and begun to hone a vision for a system of assessments in New Mexico that:

- Is applied and hands-on (i.e. performance-based, as modeled by New Hampshire)
- Affirms a collective statewide vision and local self-determination
- Measures progress and supports improvement
- Is linguistically, culturally, and real-world relevant
- Is clearly understood by parents and community
- Assesses the depth and breadth of student learning
- Gives students multiple ways to demonstrate what they know
- Prioritizes student learning

The roadmap for a new assessment system in New Mexico is under construction and building momentum. Future Focused Education is partnering with others to gather input, advocate for policy change, and re-empower teachers and communities. Students, educators, legislators, and community members are speaking up with us, and we invite you to join us.

Onward!



For more information or to partner with Future Focused Education on a stakeholder convening in your area, contact:



**FUTURE
FOCUSED
EDUCATION**

Tony Monfiletto, Executive Director
tony@futurefocusededucation.org
505-573-4024

“It’s time for change. Traditional and innovative performance assessments better measure the desired milestones and attributes for college and career readiness.”

Dr. Veronica C. Garcia,
Santa Fe Public Schools Superintendent

Introduction

In American democracy, we value the knowledge and diversity of local communities.⁴ Yet in the education arena, a long history of federal and state policies has removed significant decision-making authority from local communities, schools, and teaching professionals and placed it in the hands of the state and federal government. Indigenous communities have suffered the most egregious examples of educational disempowerment. As a consequence, the desire for self-determination runs deep in New Mexico, and provides a foundation on which we might remodel our state's approach to education and student assessment.

An ongoing devaluation of local wisdom has been one of the unintended consequences of high-stakes, standardized testing, which has played a central role in New Mexico's K-12 education system since the 2001 reauthorization of the Elementary and Secondary Education Act—known as “No Child Left Behind” (NCLB). Significant decision-making authority is consolidated in Santa Fe and Washington, far from many of New Mexico's communities.

Furthermore, the standardized tests currently used in the state accountability system measure some but not all of what communities consider important, and are limited by their format, which largely omits real-world applications of student knowledge. The tests have also meaningfully shaped curriculum and instruction to focus on basic reading and math proficiency at the expense of a broader and deeper set of skills students will need for success.^{5, 6}

Could changes to our state's assessment system now serve as a lever for re-empowering communities and educators, and lead to transformational changes throughout our education system?

New Mexico is poised to join other states nationwide in building an assessment environment that is relevant and engaging to students, reflects local context, assesses deeper learning and a broader set of competencies, responds to community and industry demands, provides meaningful data to diverse stakeholders, values the professional judgment of educators, and prepares young people for a dynamic and uncertain future. Several developments in the educational landscape make this a ripe time to explore how our state might reform its system of assessments:

First, over the past decade, we have seen numerous bright spots emerge where groups of New Mexicans are creating and innovating new approaches to assessment *in spite of* the pressures of a top-down accountability model built on standardized testing.⁷

⁴ Dewey, J. (2012). *The public and its problems: An essay in political inquiry*. University Park, PA: The Pennsylvania State University Press.

⁵ Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, 36(5), 258–267.

⁶ Achieve (2018, March). What gets tested gets taught: cautions for using college admissions tests in state accountability systems.

Retrieved from <http://www.newmexicansformichelle.com/policy/a-new-direction-for-new-mexico-s-public-schools>

⁷ Learning Alliance New Mexico (n.d.). *Alternate assessments in New Mexico: Responding to community*

Retrieved from http://learningalliancennm.org/wp-content/uploads/2016/09/Learning_Alliance_LESC_report_Nov2017_Final.pdf

Second, the Every Student Succeeds Act (ESSA) replaced NCLB in 2015, and while the law is still mainly prescriptive in regard to assessment requirements, there are opportunities for flexibility in assessment that, for the most part, New Mexico has not yet utilized.^{8,9}

Third, the recent *Yazzie/Martinez vs the State of New Mexico* decision, which declared that New Mexico has “violated the rights of at-risk students by failing to provide with a uniform statewide system of free public schools sufficient for their education,” sets the stage for a significant redesign of the support and programming offered through our school system, not simply adding or redistributing funding.¹⁰

Finally, the gubernatorial election promises a change in state and education leadership. Both leading candidates consider education a priority and are interested in changes to the assessment system but differ on the scope of those changes.¹¹

As we consider changes to New Mexico’s K-12 assessment system, the FrameWorks Institute offers an apt metaphor for our work:

*Like a general contractor, we have to remodel our educational system so that it enables our society to thrive in today’s world. ... The bad news is that remodeling creates temporary dust, noise, and inconvenience, but the good news is that when you remodel you don’t have to start from scratch — you strengthen what’s working and fix what’s not.*¹²

Taking this metaphor a step further, we must create a clear vision for what the end product of renovations to the assessment system could look like in New Mexico, but it may not be time to leap into a complete remodel. Instead, we take a long view, suggesting that our state start with initial steps toward assessment reform (i.e., one room in the house), and gradually work toward remodeling the entire building.

This paper is intended to support stakeholders in developing initial steps and a long-term vision for a new system of assessments in New Mexico. It begins with an argument for why change is necessary, followed by an exploration of foundational principles, assessment types, and systems of assessment. We then provide descriptions of important arenas for overcoming barriers to systemic change, and proposals for possible footholds where New Mexicans could start taking steps toward the long-term vision. The final section briefly introduces assessment reform happening in three states across the country that that we might learn from.

⁸ Gewertz, C. (2018, April 3). ESSA offers testing flexibility. So why aren’t states using it? *Education Week*, 37(25), 21-22.

⁹ Patrick, S., Worthen, M. & Frost, D. (2017). *Redesigning systems of assessments for student-centered learning*. iNACOL.

¹⁰ Tabet-Cubero, E. (2018, August 2). The Yazzie/Martinez school sufficiency decision is about way more than money [Blog post].

Retrieved from <http://futurefocusededucation.org/2018/08/02/the-yazzie-martinez-school-sufficiency-decision-is-about-way-more-than-money/>

¹¹ Nott, R. (2018, September 1). Education a key issue for New Mexico candidates. *Santa Fe New Mexican*.

Retrieved from <http://www.santafenewmexican.com>

¹² Kendall-Taylor, N. (2009). *Orchestrating systems and remodeling reform: Reframing education reform with simplifying models*. FrameWorks Institute, p. 15.

Why Change Assessment?

One of the principal intentions behind the standardized testing movement, and the accountability systems that have accompanied it, is a good one: As a nation, we have a responsibility to ensure students—regardless of geography and demographics—are succeeding at high levels and provided an equitable opportunity to learn important educational content standards. Beginning with NCLB, it was required that student performance on standardized tests be disaggregated by race, ethnicity, gender, and other student groups.¹³

5 Consequences of High-Stakes Standardized Testing

*Assessment and accountability systems across the nation have relied heavily on summative, end-of-year, high-stakes, standardized tests. These tests have helped illuminate opportunity and achievement gaps; supported decisionmakers in understanding how students are doing collectively; and allowed for a more strategic targeting of resources and interventions. The use of the tests, however, has had multiple unintended consequences.*¹⁴

1. In many classrooms, curriculum and instruction have narrowed to focus on tested subjects and skills at the expense of a more well-rounded educational experience. Meanwhile, with increases in automation and globalization, employers and higher education institutions are finding graduates deficient in key 21st century skills like problem-solving and communication.
2. Test scores are used as a primary measure of school quality but provide an incomplete picture of the opportunities and supports offered by a school, resulting in the public relying on incomplete information.
3. Many teachers have lost autonomy, work under undue pressure, and feel devalued, contributing to low morale and teacher shortages. When test results contribute significantly to teacher evaluation and licensure, as is true in New Mexico, it can also de-incentivize some teachers from working in high-need schools where students are less likely to score well.
4. Significant classroom time is lost to teaching test-taking skills and to administering the tests.
5. Some schools are giving more attention to “bubble” students—those on the cusp of testing as proficient—than on students who are less likely to raise the school’s scores.

¹³ No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002); United States Department of Education (2003). Report cards: Title I, Part A: Non-regulatory guidance.

¹⁴ This list was developed drawing from a number of texts: (1) Foster, M., & Immekus, J. (2018, May 7). Deeper learning for greater equity [Blog post]. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/urban_education_reform/2018/05/deeper_learning_for_greater_equity.html (2) Patrick, S., Worthen, M., Frost, D., & Truong, N. (2018). *Current to future state: Issues and action steps for state policy to support personalized, competency-based learning*. iNACOL. (3) Evans, C. M., Rose, A. S., & Leather, P. (2018, June). *Innovative assessment and accountability systems that support continuous improvement under ESSA: Practical considerations and early research*. Symposium session presented at CCSSO's National Conference on Student Assessment in San Diego, CA. (4) Zhao, Y. (2016). *Counting what counts: Reframing education outcomes*. Bloomington, IN: Solution Tree Press.

Importantly, this mandate helped illuminate gaps in opportunities and achievement for different student groups. As Kevin Welner and William Mathis articulated during the ESSA reauthorization process, however, “testing will document this, but it will do nothing to change it.”¹⁵ Instead, because standardized tests assess a narrow skillset, instructional approaches in schools and classrooms across the country have too often reflected that limited definition of what student success looks like.¹⁶

Critically, the heavy role standardized tests play in students’ experience, teacher evaluations, and the state’s accountability system has contributed to public discontent with the tests across New Mexico. Between August 2016 and March 2017, Learning Alliance New Mexico (LANM) led an effort to gather input from more than 4,000 stakeholders statewide on important educational topics, including student assessment. Unsurprisingly, “participants overwhelmingly agreed that bubble tests and annual tests are the least helpful” for parents, students, and educators, and advocated that the state should reduce testing and use instead more hands-on work and a variety of approaches to individualize assessment.¹⁷

As conceived of under NCLB and now under ESSA, the purpose of the statewide annual standardized tests has been to monitor progress on educational outcomes at school, district, and state levels, and to inform policy decisions and resource allocation. It has not been to provide actionable information to students, parents, and educators. Given the significant public investments of time and money in these tests, and the impact on pedagogy, it is time for the utility of these assessments for all parties to be considered. This is especially true for underserved students and communities and those most greatly impacted by assessment policy. We envision a new system of assessment for New Mexico where students, parents, teachers, and schools are supported first and foremost, with the assessment information needs of policymakers considered as an important, but secondary priority.

A coalition is building in New Mexico to co-create a vision for a new assessment system. Good work is already underway. Meanwhile, similar efforts are happening elsewhere in the country as states capitalize on flexibility allowed under ESSA. A clear and influential vision for assessment reform could serve as a fulcrum for changes throughout the educational system (see Appendix A for a proposed theory of change for our state), including:

- Redefining student success,
- Building teacher capacity and empowering professionalism,
- Increasing school responsiveness to local context,
- Transitioning to competency-based education,
- Rethinking school accountability, and
- Resulting in changes to school design and classroom practice.

Change will not, and should not, happen overnight. It will require that we take a long view and ask ourselves: What initial steps might we take now to move New Mexico towards a remodel of our system of assessments, and ultimately improve learning for all students?

¹⁵ Welner, K. G., & Mathis, W. J. (2015, February 12). *Reauthorization of the Elementary and Secondary Education Act: Time to move beyond test-focused policies*. Retrieved from <http://www.newmexicansformichelle.com/policy/a-new-direction-for-new-mexico-s-public-schools>

¹⁶ Conley, D.T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

¹⁷ Damle, R., Heredia-Griego, M., & Castillo, D. (2017). *ESSA Phase II Stakeholder Feedback Report*. Albuquerque, NM: Center for Education Policy Research.

What Does a New Vision Look Like?

This paper does not prescribe a vision for New Mexico’s system of assessments. Stakeholder engagement will be essential in crafting a shared vision for what we want our children to learn, know, and be able to do, and how we might assess it in more authentic and meaningful ways. This section begins by proposing a set of principles that have emerged as important among New Mexicans and others engaged in this issue across the country. A shared set of principles can help ground us in what matters as we develop a collective vision. We then go on to briefly outline different forms of assessment and discuss what is meant by a system of assessments.

What principles might guide a new assessment system in New Mexico?

Based on national research, feedback sessions conducted by LANM, and the author’s own observations and conversations with a variety of stakeholders, the following box contains a proposed list of principles that might guide reforms to our state’s system of assessments. The list is by no means comprehensive but intends to start a conversation about shared values that might inform a statewide vision.

Proposed Principles

New Mexico’s assessment system should reflect:

- ☐ High expectations for all students
- ☐ Authentic application of knowledge and skill
- ☐ Student agency and personalized learning as goals of assessment
- ☐ Student diversity of interests and talents (students do not and should not excel in identical ways)
- ☐ Educator empowerment and collaboration to design, innovate, and evaluate
- ☐ Accessibility, usefulness, and timeliness of results for all users
- ☐ Alignment with curriculum and instruction
- ☐ Responsiveness to community and industry demands
- ☐ Equity of resources and opportunities
- ☐ Inclusive assessment designs
- ☐ Self-determination of local and indigenous communities

What kinds of assessment are there?

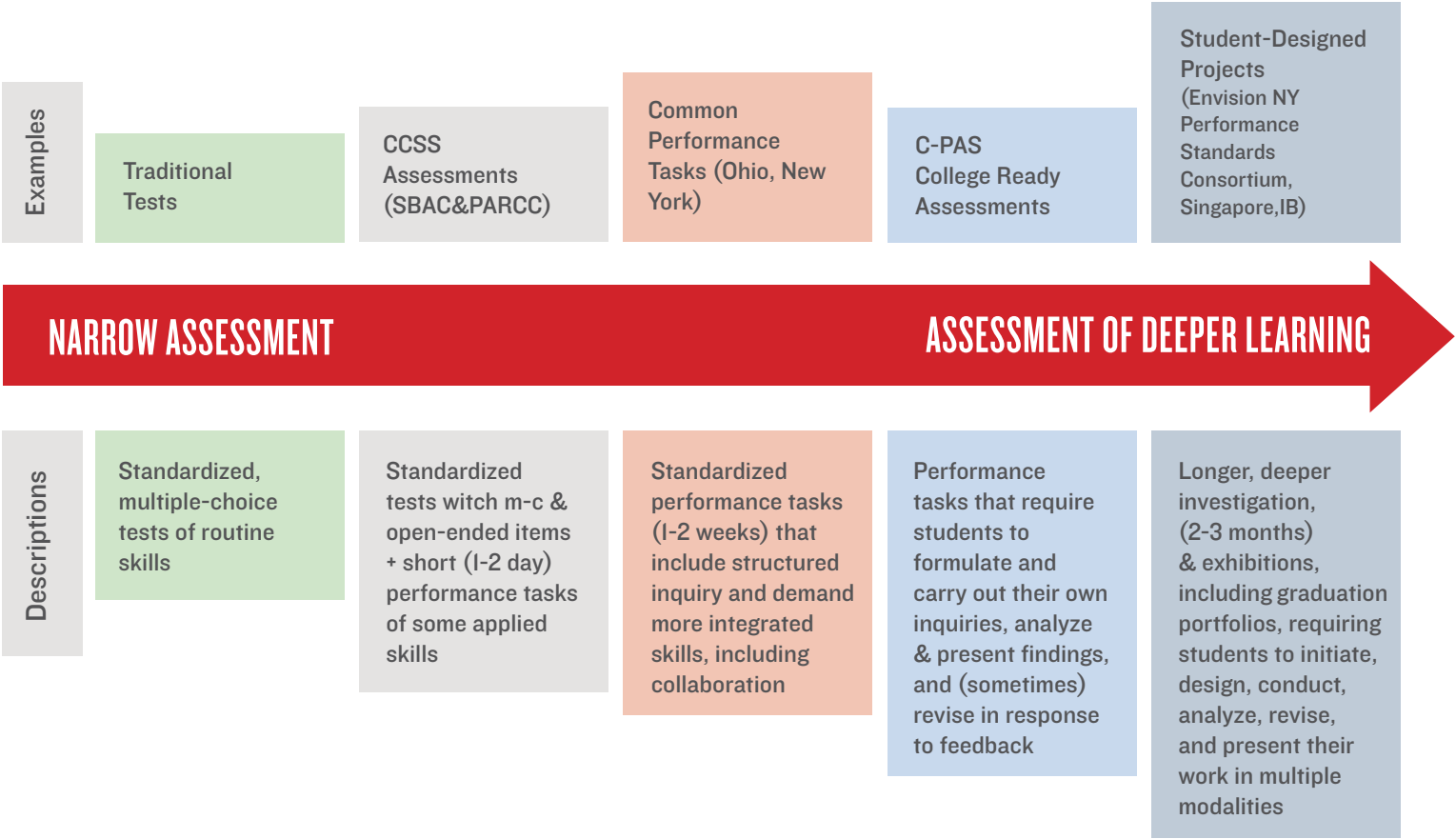
Large-scale standardized tests with primarily multiple-choice items are just one method for states to evaluate student, school, and district performance.

In the diagram below, Linda Darling-Hammond places several examples of assessment types on a continuum from those that measure more limited pieces of student learning to those that measure a breadth of cognitive and noncognitive skills.¹⁸ The three types on the right-hand side of the continuum are examples of performance assessments. Performance assessment serves as an umbrella term for a variety of formats (see the table below). Scott Marion and Katie Buckley provide the following definition:

*Performance assessments are generally multi-step activities ranging from quite unstructured to fairly structured. The key feature of such assessments is that students are asked to produce a product or carry out a performance ... that is scored according to pre-specified criteria, typically contained in a scoring guide or rubric.... [P]erformance assessment can range from 15-20 minute tasks (i.e., quite short) to multi-day activities with many scoreable units.*¹⁹

An additional approach to assessment is portfolio-based. In this type of assessment, students curate a set of artifacts from multiple projects, tasks, and assignments, and submit these for evaluation, sometimes including a culminating presentation and defense.

Assessment Continuum²⁰



¹⁸ Conley, D. T. (2015). A new era for educational assessment. *Education Policy Analysis Archives*, 23 (8).

¹⁹ Quoted on pages 5-6 in Marion, S., & Leather, P. (2015). Assessment and accountability to support meaningful learning. *Education Policy Analysis Archives*, 23(9).

²⁰ Darling-Hammond, L. (2017). *Developing and measuring higher order skills: Models for state performance assessment systems*. Palo Alto, CA: Learning Policy Institute and Council of Chief State School Officers.

The table below outlines different types of performance assessment, drawing primarily from work done by David Conley and Linda Darling-Hammond, with some extrapolation by the author of this paper.²¹

Types of Performance Assessment

TYPE OF PERFORMANCE ASSESSMENT	DESCRIPTION	EXAMPLE
Short in-class performance task	Teachers give students a task to complete that is closely tied to what is being learned in class and requires “only modest extrapolation and application.” (Conley & Darling-Hammond, 2013, p. 22)	Students analyze which cell phone plan is the best fit for them and their families based on usage patterns.
Multiday performance task	These tasks might be developed by an individual teacher but are increasingly developed in common across classrooms, schools, districts, or even across states. They often involve multiple steps: online research or an examination of provided information, and developing a product or carrying out a performance.	Students are provided with the scenario of a homeowner who is trying reduce her heating bills, and with a set of artifacts to consider (e.g. bills, temperature records). They must create a plan for the homeowner, and a pamphlet for gas company customers.
Project	Students decide what to focus on, within a provided set of parameters—often multi-disciplinary—and then complete the project over the course of several weeks or months. Teachers must provide the students with clear requirements they must meet (e.g. a scoring guide or rubric). A teacher may score a project individually, or the project may be scored by multiple teachers in a moderated process to calibrate their scores.	Following a natural disaster, students must write a paper on the government’s role, draft a response plan, make a speech to motivate citizen involvement, and prepare a poster on the environmental impact.

²¹ Conley, D.T., & Darling-Hammond, L. (2013). Creating systems of assessment for deeper learning. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Culminating Project

A student applies learning from across subject areas to one topic and generally must prepare a culminating paper or product and a presentation followed by questions from the audience—like a dissertation defense. The goal is to measure a student’s cumulative skill and knowledge through an exhibition of learning to a panel of teachers, students, and community experts.

Over the course of the year, a student participates in an internship and must prepare a demonstration of learning that meets criteria specified in a rubric.

What is a system of assessments?

Over the years, educators and policymakers have used different methods of assessment at different levels and for different purposes. However, “a collection of assessments does not entail a system any more than a pile of bricks constitutes a house,” writes Theodore Coladarci.²² In other words, a system of assessments is not simply a collection of different assessments administered to students over the course of a school year. Rather, a balanced system creates a coherent assessment experience from the student-learning perspective while serving the multiple information needs of students, teachers, administrators, parents, and policy makers.

The National Research Council’s 2001 publication, *Knowing What Students Know: The Science and Design of Educational Assessment*, has become a foundational piece in the field of assessment reform. The authors conclude that no one assessment should be designed to meet all purposes, and that

*A vision for the future is that assessments at all levels—from classroom to state—will work together in a system that is comprehensive, coherent, and continuous. In such a system, assessments would provide a variety of evidence to support educational decision making. Assessment at all levels would be linked back to the same underlying model of student learning and would provide indications of student growth over time.*²³

Nearly two decades later, there are few if any systems of assessment that exemplify this kind of alignment. In New Mexico, day-to-day instruction and classroom assessments can feel disconnected from the state-mandated End of Course Exams (EOCs) and the Partnership of Readiness for College and Career (PARCC)—both of which are used in the state’s systems for teacher evaluation and school accountability.

²² T. Coladarci. (2002). Is it a house...or a pile of bricks? Important features of a local assessment system. *Phi Delta Kappan*, 83(10), 772–774, p. 773.

²³ National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academies Press.

The high stakes placed on these state assessments can lead us to forget the state assessments are just one part of a larger assessment system that includes classroom assessments—and almost always school and/or district assessments too. In a balanced system, a variety of different types of assessment are used for different functions to meet the needs of all parties,²⁴ without the state assessments having outsized influence as they do currently in New Mexico despite the fact that they do not serve all stakeholders well. The table below identifies the purposes assessments across a system can serve for different stakeholders.

Assessments Can Provide Information to Users for Diverse Purposes

USER	LEVEL(S) OF INFORMATION NEEDED	PURPOSE(S)
Students and Parents	Individual	Support a student’s learning and understand his or her progress toward learning goals.
Teachers	Individual and classroom	Support a student’s learning and understand his or her progress toward learning goals.
School and District Leaders	School and district	Understand schools’ progress toward learning goals, support continuous improvement processes, and inform decision-making about instruction, curriculum, resource allocation, and programming.
Policymakers	District and state	Understand how students and certain student subgroups are doing overall; measure school and district quality.
Higher Education and Employers	Individual	Communicate a graduate’s competence, strengths, and areas for growth to help determine their qualifications for enrollment and employment.

²⁴ Ibid.

In a balanced system of assessments, horizontal and vertical coherence is important. Horizontal coherence means that from a student's perspective, the different assessments they take over their tenure in the K-12 system make sense together, are aligned with standards and instruction, involve them more over time as active agents in their own learning and assessment, and contribute to their learning. Vertical coherence means that from a bird's eye view, assessments from classroom to state levels meet the needs of all users, and are aligned to a shared vision, goals, and definition of student competence.²⁵

What Are the Barriers to Change?

Systemic change can take place in a variety of distinct but interrelated arenas. Here we examine four arenas where New Mexico will need to address barriers in order to make effective changes in our system of assessment.

Arena 1: Federal Policy

At the federal level, ESSA requires that states use the same test statewide to assess all students yearly in math and English/language arts from third through eighth grades, and once in high school. The law's Section 1204, however, outlines the Innovative Assessment Demonstration Authority (IADA), which waives the same-test-statewide requirement so an approved state can pilot innovative assessments in a subset of districts and use the results in their state's accountability system, while the state's other districts use the status-quo assessment. IADA allows for up to seven states, or groups of states, to participate (following the application and approval process). To date, only New Hampshire and Louisiana are approved under this authority, and the process is open for additional states to apply. Applications for this round are due December 17, 2018, and additional rounds are anticipated.²⁶

A key decision point in developing a statewide vision for assessment is whether to:

1. Apply under IADA, which would allow New Mexico to pilot a new assessment in a subset of districts and, in those districts, supplant the statewide assessment, eventually scaling statewide; or
2. Supplement the statewide assessment with innovative assessment types that would not carry any weight in the state accountability system, such as piloting new kinds of assessments in in career technical education, social studies, or the arts.

²⁵ National Research Council. (2006). *Systems for state science assessment*. Washington, DC: National Academies Press.

²⁶ U.S. Department of Education. *Innovative Assessment Demonstration Authority (IADA)* [Webpage]. Retrieved from <https://www2.ed.gov/admins/lead/account/iada/index.html>

Arena 2: State Policy

A number of state policies create challenges that restrict districts and schools from employing innovative approaches to assessment. New Mexico requires that students take the PARCC in all grades third through eleventh, which is more than what is required by ESSA. Students also take standardized End-of-Course exams (EOCs) required by state statute. Starting in grade 4, the EOCs are required in most grades and subjects.

Advocating for enabling policies that support innovation could go a long way toward building momentum and growing our state's pool of knowledge on other forms of assessment. The state might consider, for example, reducing the amount of state-mandated testing to only what is required by ESSA, thereby providing schools and districts more latitude. Another enabling policy would be to create innovation zones, also called districts of innovation. Innovation zones are an appropriate step for states "interested in shifting their role from enforcing compliance to one of supporting innovation and building capacity in districts."²⁷ Typically, an innovation zone policy offers waivers or exemptions that allow for greater flexibility to implement new models of teaching, learning, and assessment for piloting districts.

For example, Colorado passed legislation in 2008 allowing the state to designate "districts of innovation." Through this action, the state opened a door for districts to apply for greater flexibility, and waived statutes or rules specified in the district's submitted plan that would hinder their ability to implement an innovative design.^{28,29} The goal with these types of policies is that the state might learn from pilots and innovations being tested in a few districts to inform policy statewide.

David Conley and Linda Darling-Hammond point out that the PARCC measures some but not all of the Common Core State Standards (CCSS).³⁰ In addition, employers, higher education leaders, and community members nationally and locally have identified a broader range of competencies necessary for success in college, career, and life, than those in the CCSS.^{31,32} In other words, even while continuing to administer current state assessments, there is an opportunity for New Mexico to pilot supplementary assessments for learning. Performance assessments, for example, can validly measure deeper learning by asking students to apply their knowledge to real-world contexts, and can help educators, students, and families understand the importance of a more comprehensive set of competencies than those measured by the PARCC.

²⁷ Patrick, S., & Gentz, S. (2016). *Innovation zones: Creating policy flexibility for personalized learning*. iNACOL, p.1.

²⁸ Ibid.

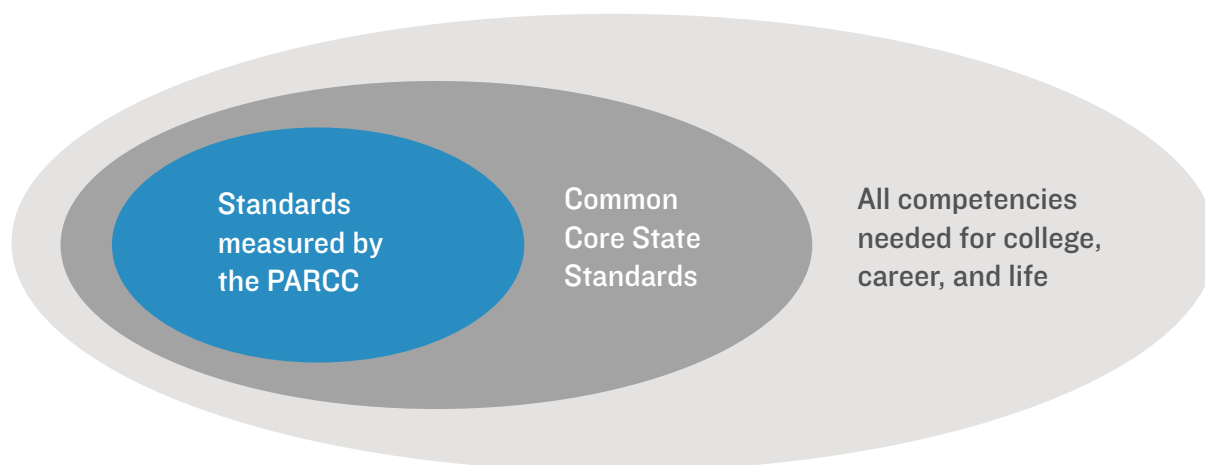
²⁹ Innovation Schools Act of 2008, § C.R.S. 22-32.5-101 (2008)..Retrieved from <https://www.cde.state.co.us>

³⁰ Conley & Darling-Hammond, 2013.

³¹ Ibid.

³² Good, D. (2017). *Graduate Profile User Guide*. Albuquerque, NM: Mission: Graduate. Retrieved from <https://missiongraduatenm.org/graduate-profile-user-guide>

PARCC Measures a Limited Set of Competencies³³



Arena 3: Culture of Accountability

Accountability systems are not a focus of this paper. The heavy reliance on standardized test scores in New Mexico's school grading system, however, means the issues of assessment and accountability are deeply intertwined. Accountability for schools and teachers is essential in any state concerned with ensuring that all students receive a quality education. The current system, however, uses top-down mechanisms with disciplinary consequences, and has received significant criticism around the state.³⁴ The approach has served to proliferate a culture of antagonism and fear among educators and educational leaders.

Local and national leaders are calling for a shift in the culture of school accountability, advocating for systems that engender collaboration and continuous improvement. Accountability as continuous improvement involves acknowledging differences in support and capacity across districts and schools, and acting to strengthen capacity for those that are struggling,³⁵ or said another way: "Accountability as continuous improvement recognizes that each student, teacher and school is in a different place on their path to meeting high expectations and that each one has room to improve."³⁶

Another helpful concept in rethinking school accountability is Richard Elmore's reciprocal accountability. He writes:

Accountability must be a reciprocal process. For every increment of performance I demand from you, I have an equal responsibility to provide you with the capacity to meet that expectation. Likewise, for every investment you make in my skill and knowledge, I have a reciprocal responsibility to demonstrate some new increment in performance. This is the principle of "reciprocity of accountability for capacity."³⁷

³³ Adapted from a diagram in Conley & Darling-Hammond, 2013.

³⁴ Perea, S. (2018, August 17). Over one-third of NM schools get D, F grades. *Albuquerque Journal*. Retrieved from <https://www.abqjournal.com>

³⁵ XQ Institute (2018). *High school & the future of work: A guide for policymakers*.

³⁶ Patrick, S., Worthen, M., Frost, D., & Truong, N., 2018, p. 21.

³⁷ Elmore, R.F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development education*. Albert Shanker Institute, p. 5.

Vertical coherence in a system of assessments supports this kind of accountability, where different parties are holding themselves and each other accountable within a system that uses assessments to fuel improvement.

Arena 4: Educator Capacity

Other than standardized tests, many forms of assessment rely on teachers' ability to design and administer, and then to evaluate student performance. Valid and reliable determinations of student mastery are essential for assessment systems that work. How do we empower teachers and build their capacity to assess a variety of evidence—not just answers on a test?

While some school leaders and teachers are thinking outside the box, pursuing unique training opportunities, and implementing non-traditional and performance assessments, they remain the exception, not the norm. Most educator preparation programs teach traditional approaches to assessment, and there is room for growth nationwide in teachers' level of assessment literacy.³⁸ Assessment literacy is “the knowledge and skills associated with designing, selecting, interpreting, and using high-quality assessments to improve student learning and to serve other important educational and policy purposes.”³⁹

Building teachers' assessment literacy and their ownership and capacity, with expert guidance and support, is key to sustainability of any assessment system. This might mean empowering teachers to look at student work in teams, across districts or schools, leveraging their collective expertise. It will also mean providing professional development, including enlisting them in determining what should be assessed and how it should be assessed, and in creating clear and calibrated rubrics to support teachers in evaluating student performance. Finally, supporting teachers in using assessments for improving teaching and learning should be a priority but is too often overshadowed by the monitoring, evaluation, and accountability purposes elevated by state and federal policy.⁴⁰

³⁸ Patrick, S., Worthen, M., Frost, D., & Truong, N., 2018.

³⁹ Marion, S., Shepard, L. (2017, June 15). *Assessment literacy to support competency-based education systems and other deeper learning efforts* [Powerpoint], p.2. Retrieved from https://www.inacol.org/wp-content/uploads/2017/06/Assessment-Literacy-for-CBE_061417_compressed.pdf

⁴⁰ Marion, S., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2018). A tricky balance: The challenges and opportunities of balanced systems of assessment. National Center for the Improvement of Educational Assessment. Retrieved from https://www.nciea.org/sites/default/files/in-line-files/A%20Tricky%20Balance_092418.pdf

Where Do We Start?

Remodeling the system with stakeholder input and engagement is essential to developing a long-term vision for changes to New Mexico's assessment system and any related changes to the accountability and teacher evaluation systems, which both lean heavily on the high-stakes standardized tests that dominate our current system. A sudden and complete overhaul is neither likely nor advisable, given the complexity and interconnectedness of the state's assessment, accountability, and teacher evaluation systems, and the large investment of time, money, and learning many districts have made in adopting the PARCC just four years ago.

Instead it is time for a committed group of stakeholders to outline a plan with short- and long-term strategies for moving our state forward. What can we do now to advance New Mexico toward an assessment system that recognizes the broader and deeper set of competencies our students need for success, and better meets the needs of all users?

The shift from NCLB to ESSA could be an indication that additional flexibility will come from the federal level in the next decade. What can we do now that will position us well for additional flexibility in assessment and accountability systems?

New Mexico is home to many bright spots where good work and innovation are happening, which we can build on. Below are three possible footholds to consider in determining initial steps, and there are others.

Foothold 1: Graduate Profile

A strategy rooted in local wisdom, the graduate profile is a community-informed articulation of competencies our future graduates will need for success. There is significant evidence (from skills gap studies, employer surveys, and rates of remediation needed among first-year college students) that many school systems are not adequately preparing students for college, career, and life.⁴¹

Engaging stakeholders to redefine student success with more breadth and depth than the current emphasis on reading and math is a logical place for New Mexico to start. The resulting graduate profile can become the foundation for a new system of assessments as well as a tool for aligning with higher education and workforce expectations. A graduate profile can also be a rallying cry that the goals we have for our young people are not solely the domain of schools and districts but rather a shared responsibility across our communities, families, businesses, and higher education institutions.⁴²

⁴¹ For example: (1) The Chronicle of Higher Education & Marketplace. (2012). *The role of higher education in career development: Employer perceptions* [Powerpoint]. Retrieved from <https://chronicle-assets.s3.amazonaws.com/5/items/biz/pdf/Employers%20Survey.pdf> (2) Dauphinee, T., Reagan, S., Plummer, J., Timm, J., McCloud, A., Hollis, J., & Dekker, J. (2015). *Albuquerque hiring skills gap study: The Albuquerque Business Education Compact (ABEC)*. Albuquerque, NM: University of New Mexico Center for Education Policy Research. (3) *Complete College America* (n.d.). *Corequisite remediation: Spanning the completion divide*. Retrieved from <http://completecollege.org/spanningthedivide/>

⁴² Patrick, S., Worthen, M. & Truong, N. (2017). *Redefining student success: Profile of a graduate*. iNACOL.

There are several examples of graduate profiles being developed in New Mexico. Santa Fe Public Schools very recently adopted “attributes of a graduate” based on input from a broad base of community and internal stakeholders. The district is now working to develop aligned curriculum and assessments.⁴³ In addition, Mission: Graduate, an initiative of the United Way of Central New Mexico, worked with Albuquerque Public Schools leadership and community partners to develop the Central New Mexico Graduate Profile, based on a careful analysis of more than 50 focus groups in 2016. The Profile and accompanying User Guide could provide a starting point for local communities to develop their own, or for the state to adopt a statewide version. See Appendix B for a copy of the Central New Mexico Graduate Profile and a link to the User Guide.⁴⁴

Foothold 2: Grades and Subjects with Less Required Testing

While introducing new assessment approaches would be possible in any course, in most cases it would mean adding assessments in addition to the PARCC and EOCs, unless New Mexico does one or both of the following:

1. Change statewide policy and practice for EOCs, including allowing for waivers of the current statewide EOC requirements, and/or
2. Applies and gains approval under Section 1204 of ESSA, which would allow New Mexico to supplant the PARCC (and potentially the EOCs) in most/all grades in a subset of districts in order to pilot a new system of assessments.

Regardless of whether the state takes these steps, identifying courses and grades with fewer or no required statewide tests could be a compelling place to start expanding the use of performance assessment and other innovative approaches. Innovating in grades and subjects with less required testing and advocating for state recognition of these practices in lieu of current requirements could be initial steps toward a long-term vision in New Mexico. In fact, innovative practices are already happening in many classrooms and schools across New Mexico.

One promising example has been codified into the policies of the state’s Public Education Department: the State Seal of Bilingualism-Biliteracy. Students can earn the seal in multiple ways, including a portfolio process described as including “a presentation, an interview with a panel composed of three or more members of the district’s education staff and community who are proficient in the target language other than English, and a student-produced work sample, written when appropriate.”⁴⁵

⁴³ V. Garcia, personal communication, October 14, 2018.

⁴⁴ Good, D., 2017.

⁴⁵ New Mexico Public Education Department. State seal of bilingualism-biliteracy [Webpage]. Retrieved from <https://webnew.ped.state.nm.us/bureaus/languageandculture/seal-of-bilingualism-biliteracy/>

In addition, New Mexico's adoption of the Next Generation Science Standards is a promising step towards a greater use of performance assessment in science, replacing the current required Standards Based Assessment (SBA) administered in grades 4, 7, and 11.⁴⁶

Foothold 3: Career and Technical Education

Career and Technical Education (CTE) programs across the state are a natural fit for piloting new approaches to assessment. Many CTE programs are practiced at using performance tasks to assess progress toward learning targets, and depending on the program, are preparing students for performance-based assessments that result in an industry-recognized credential in their specific field.⁴⁷

What might it look like to pilot innovative assessments in CTE programs that evaluate the broader set of "soft" and technical skills that employers look for? The Concord Regional Technical Center in New Hampshire offers a compelling example of a CTE program that developed its own assessment system aimed at measuring a more comprehensive set of hard and soft skills. Students in the program participate in performance review process that includes meetings with their teachers and a series of performance and portfolio-based assessments.⁴⁸

Industry and higher education leaders have made clear that a broader array of competencies is important for all graduates and should not be confined to CTE programs. Piloting new assessment systems in CTE programs could then inform approaches to instruction and assessment across all subjects and courses. What's more, CTE programs might have lessons for other subjects in how student learning can be credentialed. There is significant innovation taking place in digital and micro-credentials that might have application here. For example, could credentials earned based on performance assessments in a breadth of skills provide employers and colleges with more meaningful information on a students' qualifications than their GPA and seat-time-based diploma currently do?

⁴⁶ NGSS Lead States (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

For more information: <https://www.nextgenscience.org/assessment-resources/>

⁴⁷ E. Perea, personal communication, October 5, 2018.

⁴⁸ Rothenberg, S. (2014, April). The CRTC SPUR process. *Techniques*, 50-54. Retrieved from <https://www.acteonline.org/wp-content/uploads/2018/05/Techniques-April2014-CRTCSPUR.pdf>

What National Bright Spots Can We Learn From?

New Mexico is unique in geography, demographics, history, and culture. Even so, examples of both small- and large-scale efforts at assessment reform in other states can be instructive as we envision what is possible. Here we very briefly reference three examples that could be sources of additional learning.

New Hampshire

New Hampshire's Performance Assessment for Competency Education (PACE) initiative has been in place and expanding since March 2015, when the state received permission to use local and common performance assessments for accountability purposes in place of required standardized testing in most grades and subjects. In its first year, four districts participated in PACE. That number has now expanded to more than twenty. The state's application under ESSA Section 1204 was recently approved, allowing PACE to continue and to scale in the coming years.

Virginia

The Virginia legislature removed required end-of-course exams called Standards of Learning in several grades and subjects in 2014, requiring instead that districts use local assessments, and encouraging the use of performance assessments. In addition, the state has developed a profile of a graduate through a process based in community conversations. The profile includes content knowledge, workplace skills, community engagement and civic responsibility, and career exploration. Virginia is currently working to align the state's standards and diploma requirements with the profile of a graduate.

Hawaii

In 2015, Hawaii's Board of Education adopted a set of learning outcomes known as HD, which span the K-12 system and reflect Hawaiian culture and values, including a sense of Belonging, Responsibility, Excellence, Aloha, Total Wellbeing, and Hawai'i. A grantee of the Assessment for Learning Project, Hawaii's Department of Education is now developing a culturally responsive assessment framework that aligns with the HD learning outcomes.

Conclusion

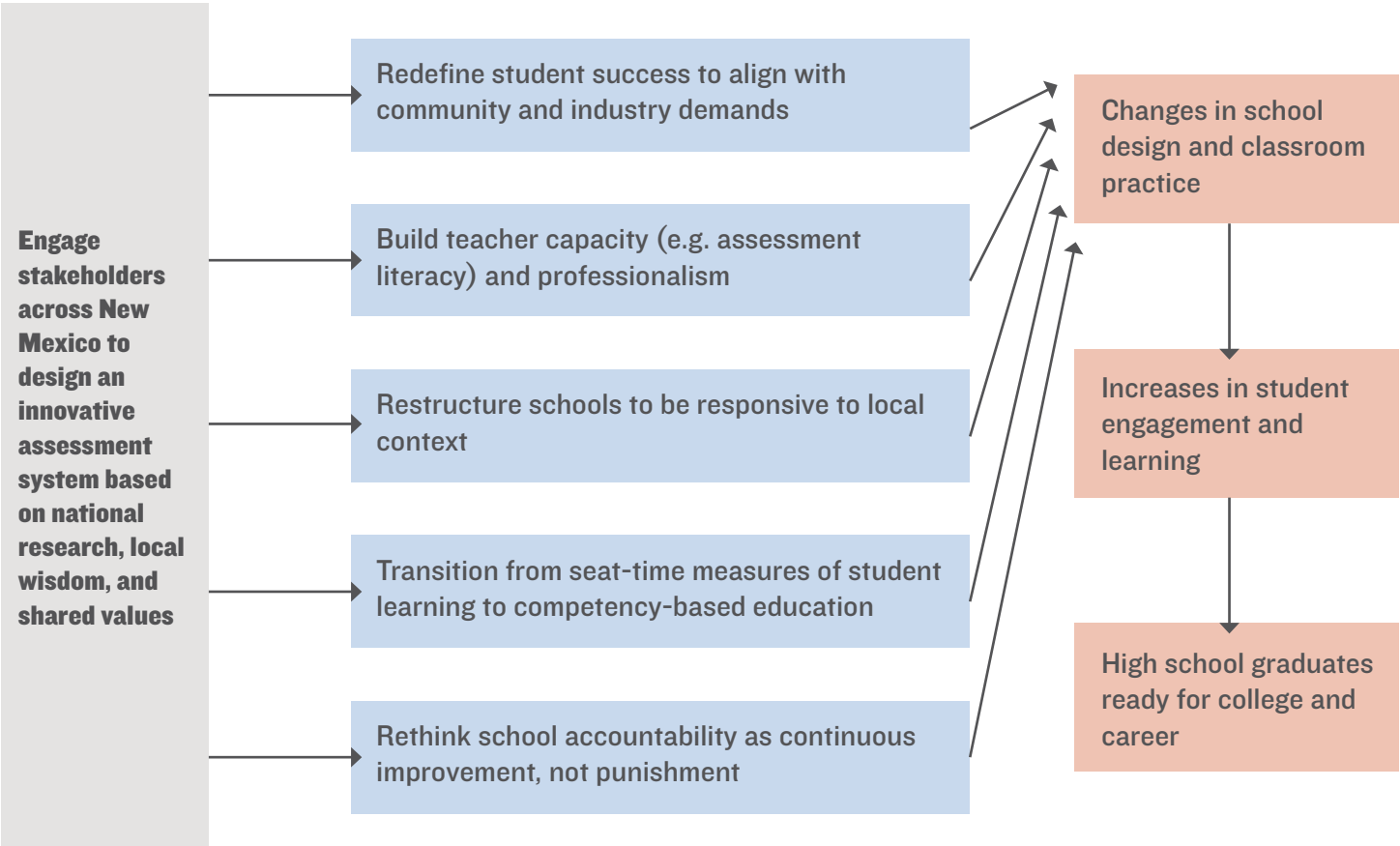
Many have called for educational reform in New Mexico, and the time is ripe for change. A new vision for assessment in our state may provide a fulcrum for transformation throughout the system. Indeed possibilities for other changes could be thwarted if assessment is not rethought because of the heavy role it plays in students' experience, teacher evaluations, and the state's accountability system.

While far from comprehensive, this paper is intended to support an engaged coalition of stakeholders in developing a new vision for assessment in New Mexico. Most likely, our vision will not mean a fast-paced and comprehensive overhaul of the current assessment system. Much like remodeling a house by starting with one room, change in New Mexico's assessment system will require a plan for initial steps, medium-term goals, and a shared long-term vision.

Let us learn from examples nationally, and build on our strengths locally. Let's get started today.

Appendix A: Proposed Theory of Change

Our theory is that changes to our state’s assessment system are a key lever for needed improvements throughout our educational system. A growing coalition of engaged stakeholders are invited to help refine this theory and collectively develop recommendations for New Mexico’s upcoming governor on an authentic and effective approach to reform.



Appendix B. Central New Mexico Graduate Profile

See <https://missiongraduatenm.org/graduate-profile> for more, including access to the User Guide, which further explains the competencies included in the image below.



Acknowledgements

With gratitude to advisors and readers:

Carrie Robin Brunder,
Zia Strategies

Veronica C. Garcia,
Santa Fe Public Schools

Susan Lyons,
Center for Assessment

Ian Esquibel,
Oak Hill Coaching and Consulting

Paul K. Leather,
Center for Innovation in Education

Susan Patrick,
iNACOL

Carla M. Evans,
Center for Assessment

Sarah Lench,
Center for Innovation in Education

Gwen Perea Warniment,
LANL Foundation

Special thanks to:

Dr. Veronica C. Garcia and Santa Fe Public Schools for joining Future Focused Education in co-hosting the stakeholder convening on October 18, 2018, for which this paper was written.

Dr. Susan Lyons and Dr. Carla Evans of the Center for Assessment for their consultation on this white paper, presentation at the October convening, and hosting a team from New Mexico on a learning tour in New Hampshire in November 2018.

