Overcoming Barriers to School-Based Large-Scale Assessment with the Support of Artificial Intelligence Tools[1]

Charlie DePascale, PhD
Psychometric Confections, LLC

Abstract

Despite decades of technological advances and calls for more authentic assessment, large-scale assessment in elementary and secondary schools in the United States remains characterized by infrequent, external, on-demand, standardized tests consisting primarily of selected-response items. It could even be argued that advances in technology that have facilitated the growth of computer-based and computer-adaptive testing have exacerbated the issue by being more focused on increasing efficiency than on enhancing authenticity, relevance, and utility. In many ways, the type of state-supported, school-based and curriculum-embedded authentic assessment of higher order thinking and 21st century skills envisioned in the 1990s seem farther from reality than ever before. The advent of artificial intelligence tools that can be applied to support both assessment and instruction, however, offers the promise of overcoming the barriers to state-supported, school-based, large-scale assessment.

In this presentation, I begin with a historical review of the calls for more authentic assessment in the late 1980s that ushered in a period that might be described as a golden age of innovation in large-scale assessment in the early- to mid-1990s. All that glitters was not gold, however, and nearly all those ambitious attempts at large-scale, school-based, performance assessment were short-lived as they were unable to overcome a wide variety of barriers to their successful implementation and use. In the second section of the presentation, I identify and discuss critical barriers such as psychometric and measurement challenges, logistical issues, and concerns related to efficiency, capacity, cost, control, and infrastructure. The relative importance, strength, and persistence of each barrier initially and over time is analyzed to paint a picture of the current assessment, instructional, and accountability landscape. The final section of the presentation is devoted to a discussion of emerging artificial intelligence tools and the ways in which they can be applied overcome persistent barriers related to ongoing data collection, psychometrics, scoring, and most importantly, instructional infrastructure. The call is for assessment developers, policymakers, and educators not to re-imagine large-scale

[1] Draft prepared for presentation at the 8th International Association for Innovations in Educational Assessment Conference. November 5, 2025.

assessment, but to revisit and recommit to the vision of state-supported, school-based, authentic assessment that has been gathering dust for nearly forty years.

Introduction

In the oft-quoted aphorism, the philosopher George Santayana warns "those who cannot remember the past are condemned to repeat it." As I imagine the future of state testing, I call on the field to both remember *and* repeat a period from our not-so-distant past. Specifically, I hope that the field can replicate the bold vision, hope, sense of purpose and possibility, and innovativeness of the early 1990s, a period that I have referred to as the Golden Age of PK-12 state testing in the United States (DePascale, 2016). At the same time, taking Santayana's words to heart, I hope that the field remembers the barriers that prevented us from implementing that bold assessment vision at that time. Because in my heart of hearts I think that we had identified the right assessment solution in the 1990s, but in a very real sense were ahead of our time and our capacity to implement that solution. Now, with the clarity of hindsight, advances in technology, and the support of rapidly emerging artificial intelligence tools I am confident that we can finally efficiently and effectively implement the vision of school-based state-supported assessment at scale.

To accomplish that goal, however, we do have to clearly remember and deeply understand the barriers to implementation that thwarted our original efforts. For those of us directly involved in large-scale testing whether as psychometricians or assessment specialists, the barriers foremost in our minds are likely technical. To be fair, there is good reason for that. There was a plethora of well-documented measurement challenges that we failed to overcome at the time, including issues within each of the three foundational areas of educational measurement and testing: validity, reliability, and fairness (Kirst et al, 1996; Koretz, 1992; Koretz, 1995; Koretz et al, 1998). All of which contributed to concerns about the comparability of results at both the individual student and aggregate levels. In reality, those measurement challenges, however daunting, paled in comparison to the technical and human infrastructure barriers associated with the transition from external, on-demand standardized tests consisting of selected-response items to student assessment comprising state-administered constructed-response items and direct writing tasks along with school-based portfolios and complex performance tasks. Although many of the technical infrastructure challenges have been addressed successfully due to the many technological advances that have occurred since the 1990s, human infrastructure to support school-based assessment in PK-12 schools remains a significant barrier and one that has been largely unaddressed since the 1990s even as the amount of testing and the stakes associated with testing increased dramatically. Fortunately, I believe that

strengthening human infrastructure is an area in which AI tools can have the greatest impact in support of state-sponsored, school-based student assessment.

The vision of the future of large-scale PK-12 assessment in the United States has been clear since the early 1990s. Standardized tests consisting of selected-response items, long the staple of large-scale testing, would no longer suffice. Those external, on-demand tests, most often measuring recall and other low-level cognitive skills would need to give way to more authentic forms of student assessment designed to measure higher-level cognitive skills such as critical thinking and problem solving as well as an emerging set of skills deemed necessary for postsecondary success. Why then, decades later is state testing still dominated by external, on-demand tests?

To move forward successfully, we must first look back.

The 1980s: Excellence, Authentic Assessment, and the Backlash Against Norm-Referenced Testing

In 1988, an article published by J.J. Cannell, a physician from West Virginia, sparked the backlash against traditional, standardized norm-referenced tests and testing that, at a minimum, accelerated the shift to criterion-referenced, standards-based testing in the United States. Dubbed The Lake Wobegon Effect, after a fictional town in Minnesota made famous by storyteller Garrisson Keillor where "*all the children are above average*" Cannell questioned the validity of results showed that all 50 states in the United States were performing above the *national average*.

At the same time, educators such as Grant Wiggins were citing the need for more authentic forms of assessment of student performance; that is, tests that more closely resembled the knowledge and skills that students were expected to possess and the tasks that they were expected to able to perform (Wiggins, 1990). Consistent with an argument still heard today, such tests would also be able to provide better feedback to teachers to guide instruction and support student learning.

Both the backlash against norm-referenced testing and the call for more authentic assessment occurred in the wake of the release of the 1984 report, *A Nation At Risk*, which questioned the quality of public schooling in the United States and had the effect of adding *excellence* to the federal government's longstanding focus on *equity* in public education. It is a relatively straight line from the publication of A Nation at Risk to the 1998 Education Summit, Goals 2000, the Improving America's Schools Act of 1994 (which mandated the establishment of state content standards and tests assessing those standards), and finally the passage of No Child Left Behind in 2002, adding strict accountability regulations to the

requirements for state standards and tests established under IASA as well as dramatically increasing the amount of required state testing.

It can be argued that NCLB effectively ended any serious consideration of or work on state-supported, school-based assessment. The sheer increase in the volume of state testing required under NCLB, including additional testing requirements for students with disabilities resulting from Individuals with Disabilities Act (IDEA) and the Americans with Disabilities Act (ADA), was enough to strain the capacity and resources of most state departments of education. Continued research and development into innovative school-based forms of authentic assessment were set aside.

It is the work that took place in state testing between the publication of The Lake Wobegon Effect in 1988 and the passage of NCLB in 2002 that established the foundation for the type of future of large-scale state testing envisioned in this paper; that is state-supported, school-based student assessment supported by artificial intelligence tools.

The 1990s: A Golden Age of State-supported, School-Based Assessment

Buoyed by the appetite for education reform and authentic assessment, a handful of states, organizations, and testing companies launched innovative assessment programs intended to change state testing fundamentally, and in fact, redefine the role of the state in state assessment. Notable innovative assessment programs from that era include:

- Kentucky Instructional Results Information System (KIRIS)
- Vermont Portfolio Assessment Program
- California Learning Assessment System (CLAS)
- Maryland School Performance Assessment Program (MSPAP)
- The New Standards Project

Each program expanded state testing beyond traditional selected-response (i.e., multiple-choice) items to more authentic forms of student assessment including direct writing samples, constructed-response items, and other performance-based items and tasks. Each of those programs, to some extent, shifted the locus of control of state assessment from the state to the schools.

A decade later, however, it was clear that all that glitters was not gold. The Golden Age was at best gilded, and at worst, tarnished beyond repair. KIRIS, CLAS, and the Vermont Portfolio Assessment Program were distant memories. The assessment portion of the New Standards Project was a shell of its original grand vision, reduced to a small set of on-demand “performance tasks” administered with and linked through a traditional norm-referenced standardized test. MSPAP may have survived the longest, but eventually it too

succumbed to the federal requirements of No Child Left Behind (NCLB) to administer the same reading and mathematics tests to all students and to annually test and report individual students results for all students in grades 3 through 8 and at least once in high school.

Whether it would have been possible for these innovative assessment programs to recover from their initial shortcomings remains an unanswered, unanswerable, and largely unasked question as the demands of NCLB and test-based accountability moved state testing in another direction.

Barriers to School-Based Assessment

From the beginning, there were three major categories of barriers to school-based assessment: measurement challenges to the validity of test results; technical infrastructure, and human infrastructure.

As noted above, the first of these, measurement challenges to the validity of test results were well-documented and widely publicized. To a large extent, they were also well-founded. From a measurement perspective, the field moved too quickly, got out over its skis, or as expressed in the highly descriptive metaphor popularized at the time, was flying the plane while building it (Rothman, 1995)[2]. In all honesty, not only were we flying the plane while building it, we were struggling also to design and find suitable parts for our assessment plane.

The measurement challenges, however, were all eminently solvable with a little time and forethought. The reliability of scoring portfolios, for example, was one of the biggest criticisms of those innovative assessment programs. Reliability of scoring, however, was also a relatively easy problem to correct, and one which largely had in fact been corrected by the time the assessment programs were cancelled. The scoring problem was much more a human infrastructure problem than a measurement problem, and one of many areas in which the field tried to do too much too quickly (Hill, 2000). Barriers related to measurement issues such as generalizability, equating, and comparability although formidable are also manageable[3].

Logistical challenges in implementing, maintaining, and sustaining school-based assessment programs were very real in the 1990s, but have largely been overcome by

[2] The "flying the plane while building it" metaphor has taken on more positive connotations in recent years but was definitely a negative description in the 1990s.

[3] It may appear that I am glossing over critical measurement issues that remain challenges, but I believe that it is true that none of them are insurmountable given the appropriate design and level of expectations.

technological advances in the intervening years and the shift to computer-based testing by all states in 2015 (Moncaleano et al, 2018).

By far, the biggest barrier to the implementation of school-based student assessment in the 1990s were challenges related to supporting the human infrastructure critical to the success of such programs. What is human infrastructure? As described in a recent blog post (Digital Wisdom Collective, 2025):

> *Human Infrastructure is the network of people, relationships, knowledge flows, and norms that enable work to actually happen inside an organization. It's the living connective tissue between strategy and execution and embodies the capacity to align, collaborate, and adapt when technology changes.*
>
> *It shows up in:*
>
> - *Mid-level leaders who bridge the gap between business vision and technical reality*
> - *Shared mental models that help teams work across silos without endless meetings*
> - *Teams' confidence to spot problems and voice concerns before they derail projects*
> - *Trust networks that carry initiatives across the finish line when processes break down*
> - *Informal knowledge flows that embed wisdom and make complexity navigable*

As we stand here in late 2025, human infrastructure remains the biggest barrier and greatest challenge to implementing state-supported, school-based assessment, and the area in which AI tools offer the most potential and promise for support.

Human Infrastructure, School-Based Assessment, and the Use of AI Tools

The biggest misstep in attempting to implement state-supported, school-based assessment in the 1990s was severely underestimating the amount of time and resources that would be needed to prepare educators, the general public, and other key stakeholders for their use. That preparation requires a multi-step and multi-year process that at the highest level must include the following four stages:

1. Gaining trust and buy-in from all participants and stakeholders
2. Allowing adequate time for acclimation and use

3. Providing continuous support for implementation and use
4. Providing support for the interpretation and use of assessment results

In the earliest efforts at implementation in the 1990s, Stage 1 (gaining trust and buy-in) was immediately and irreparably undermined by positioning the innovative school-based assessments as an instrument of a high-stakes accountability system administered by the state.[4] Rather than serving as an instructional partner the state was accurately viewed as an external evaluator and the assessment viewed as an add-on and disruption to instruction rather than something to be integrated as a central component of instruction.

The key to successful implementation, therefore, does not begin with the support of AI tools, but merely sound policy and practice that plants the state-supported assessment seed in the schools and allows it to take root before using its results as part of a state accountability system. Nurturing that seed so that it grows and blooms, however, is something that can greatly benefit from the use of AI tools.

The first step in the process is to gain at least provisional buy-in from all key stakeholders, including district and school administrators, teachers and their representatives, students, and parents/guardians. The process of securing buy-in must be a collaborative one rooted in determining how best to position state-supported assessment instruments and tools to support teachers' instructional needs and promote student learning. AI tools can play a pivotal role in this phase of implementation by supporting the iterative U/X design process.

Stage 2 (allowing adequate time for acclimation and use) is critical to moving from provisional buy-in to actual buy-in. Time alone, however, will not be sufficient unless it is accompanied by the supports described in Stages 3 and 4. It is in these two stages that I envision AI tools being most helpful in overcoming barriers to the successful implementation of school-based assessment programs and the effective use of the results of those programs. AI tools can be used to provide direct support to teachers (and to students) at a level and degree that the state will never have the capacity to match. That support begins from the earliest stages of implementation and continues through interpretation and use of results.

AI tools should be able to play a key role in supporting teachers in determining when it is most appropriate to administer a particular assessment, whether that decision is being made at the curriculum level when planning scope and sequence of instruction for an

[4] A mistake that was repeated under the Every Student Succeeds Act and its Innovative Assessment Demonstration Authority (IADA). High-stakes uses of assessment is often associated with negative teacher behaviors and score inflation on such assessments (Koretz, 2008).

entire school year or semester; at the instructional level when planning lessons for a particular unit or cluster of standards; or at the individual student level when determining whether the student is ready to participate in the assessment and move on to the next unit.

Scoring and providing exemplars of student work tied directly to established benchmarks (i.e., achievement standards) or containing examples of particular points of emphasis is obviously an area in which AI tools will be able to significantly reduce capacity concerns at both the state and school level.

The area which I think holds the biggest potential for AI tools to help strengthen human infrastructure and overcome barriers to state-supported, school-based assessment is through supporting interpretation and use; that is, in assisting teachers in making sense of the test results in front of them at any given time and deciding what to do next for an entire class, a group of students, or an individual student. This piece of information has been the missing link in virtually every instructional- and assessment-based intervention or initiative designed to place more actionable information in the hands of teachers in real time.

I witnessed this state of affairs firsthand in the first school-based research project I was involved in the 1980s as a graduate student at the University of Minnesota. We were evaluating teachers' use of a computer-based system that allowed all students to respond to teacher questions via the use of a keyboard and for teachers to see and process those responses in real-time. Although we were able to help teachers overcome the logistical and mechanical barriers to implementation, we were at a loss when it came to providing support with respect to answering the question, "What do I do next?" which was inevitably the first question asked by most teachers.

The same was true in the late 1990s in the early years of the Massachusetts Comprehensive Assessment System (MCAS) where we provided teachers not only with overall scores and performance levels of individual students, but also with 100% of the test items and actual student responses to those items.

The same was true between 2004 and 2014 as we worked with educators in Rhode Island to implementation a school-based system of Proficiency-Based Graduation Requirements (PBGR) in Rhode Island high schools. The state hired a cadre of retired teachers and other content experts to support teachers, but the capacity of the cadre to provide the level of information needed by teachers in a timely manner was insufficient to meet the need and overcome the barrier to the effective use of test results.

Current state assessments, at their best, offer information about two steps in the instructional and process:

1. A target or endpoint that includes a solid description of the knowledge and skills a student attaining the goal will possess and be able to exhibit.
2. A "sense" of where the student is now relative to that target.

Some assessments may even attempt to provide some measure of the distance between a student's current location and the target – often expressing that distance in terms of scaled scores or scores converted to a time-based metric. Such efforts more often than not fall short of their desired level of utility (Betebenner and DePascale, 2024).

What has been missing from the state assessment landscape are attempts to use data gathered over time from state testing to model the learning process and optimize the path that an individual student should follow to progress from their current location to the target – a path that will differ among students. AI tools can play a vital role in supporting that type of modeling. But AI tools will only be effective if we move state testing in the right direction.

Two Roads Diverged and a Proof Point

The years between 2010 and 2015 were a period of hope, excitement, and high expectations for the future of large-scale testing in the United States (Domaleski, 2019). On the heels of the state-led development of the Common Core State Standards (NGA, 2010) and the immediate and almost universal adoption of those college readiness standards, the federal government announced the Race to The Top Assessment Program, committing $330 million to the development of "next generation" state tests aligned to those standards. In September 2010, US Secretary of Education Arne Duncan announced the two winners of the assessment competition: the Smarter Balanced Assessment Consortium (Smarter Balanced) and the Partnership for the Assessment of Readiness for College and Careers (PARCC). Like the almost universal adoption by state of the CCSS, virtually all of the 50 states belonged to one of the assessment consortia (or in some cases, belonged to both).

As noted by Sec. Duncan, both assessment consortia were poised to take large-scale state testing in the United States "beyond the bubble test." Both assessment programs remained external, on-demand, tests administered annually at the end of the school year.[5] Both assessment programs provided the catalyst necessary to finally push state testing from paper-and-pencil to computer-based testing.

Interestingly and significantly, neither consortium proposed a school-based assessment program similar to those attempted in the 1990s. On the contrary, Smarter Balanced, proposed a computer-adaptive test (CAT) design supplemented by a performance task.

[5] Although still technically end-of-year tests, administration windows have expanded to accommodate the use of computer-based testing with limited computer resources in schools.

PARCC, in contrast to Smarter Balanced, proposed an expanded version of the existing state-of-the-art custom state test design: a test consisting of a variety of item types, including constructed-response items, short-answer items, essays along with selected response items. PARCC also included new item types made feasible by the switch to computer-based testing (CBT) and a performance task.

Smarter Balanced, although innovative in its introduction of CAT to state testing, was, and remains, a relatively bare bones test, efficiently providing an estimate of a student's overall level of proficiency (i.e., achievement of grade-level standards, progress toward college-and-career readiness). Reflecting the desires of its member states, Smarter Balanced prioritized efficiency in a large-scale state summative assessment, drawing on IRT to estimate students' overall level of proficiency with an adequate level of precision along the performance continuum. In doing so, Smarter Balanced likely serves as a model for stand-alone, large-scale state summative tests for the foreseeable future.

The path that PARCC chose proved much rockier. Almost immediately, it became obvious that the simply was too big – a state testing program on steroids (DePascale, 2015). In sharp contrast to Smarter Balanced's emphasis on efficiency, PARCC was unwieldy and exerted a tremendous burden on schools to administer while offering none of school's desired benefits of state testing such as immediate (or at least timely) return of results and results that contained more detailed, actionable information about student performance to inform instructional decisions. Consequently, it can be argued that PARCC violated the implicit social contract that undergirded state testing since its inception. Educators will disrupt their instructional routine to administer state tests with fidelity and encourage motivated student performance as long as those tests can be administered easily within a reasonable amount of time and have minimal impact on them[6]. Ultimately, all of those factors led to the demise of the PARCC consortium and assessment, despite generally high regard for the quality of its items.

As educators and lifelong learners, however, we view every failure as a learning experience. The takeaway from the PARCC experience is (or should be) that as a field we have reached the limit of what we can hope to accomplish with a state assessment program consisting of external, on-demand tests administered annually at the end of the year (or even such a test administered at 2-4 regular intervals across the school year). If we seek authentic assessment of complex knowledge and skills, we need to move beyond this model of assessment. Although it has served us relatively well for decades, it is neither designed nor

[6] In addition to push back from the field because of its design, PARCC also suffered from the shift to higher college-readiness achievement standards and being introduced at a time of high-stake school accountability and an ill-conceived and short-lived attempt at federal overreach into educator evaluation.

well-suited to meet the demands that we are now placing on state assessment; that is, demands for high-quality assessment (CCSSO, 2014), demands to measure high-level cognitive skills as well as complex 21st century skills such as critical thinking, creativity, collaboration, and communication, demands to engage students, and demands to provide timely, actionable information to teachers to inform instruction.

There are some who may hold out hope that AI tools will somehow enable us to reimagine and reconfigure external, on-demand large-scale testing in a way that meets most of those demands by combining the efficiency of a CAT with the level of authenticity desired. I do not think that such an outcome is attainable. The roads to efficiency and authenticity followed by Smarter Balanced and PARCC, respectively, do not intersect. Further, those roads are not merely parallel, they are constantly diverging – growing farther apart with every advance that moves each down its chosen, well-defined, and well-travelled path. Rather, the level of efficiency and authenticity we desire can only be achieved by moving state testing in a radically different direction, a direction that we envisioned but were unable to attain in the 1990s.

Back To The Future With State Assessment

With PARCC as the proof point that we have reached the limits of external, on-demand testing, where do states turn for the future of state assessment. Where does the field turn to achieve the level of authenticity and efficiency needed. The answer today, as it was in the early 1990s, is to the schools. Authentic assessment occurs in schools with teachers and students. It occurs on a regular basis. It occurs over extended periods of time. It occurs as an integral component of and/or natural result of instruction. Although the field was not able to overcome barriers to successful implementation of school-based assessment in the 1990s, there some was evidence of positive consequences with regard to effects on teachers' instructional practices (Stecher et al, 1998).

If we do turn to schools for state assessment, why will the outcome be different this time? What role will the state play? What role will AI tools play in supporting and enhancing states' efforts?

Role of The State and the Use of AI Tools

The role of the state in this reimagined, new and improved, school-based model of assessment will be three-fold.

1. First, the state will serve as a *clearinghouse* for high-quality, school-based assessment instruments and tools. States, individually or collaboratively, will oversee the development of assessment tools and instruments aligned to their standards and will also vet and certify for use such tools and instruments produced

by others. Louisiana serves as a model of a state that has fulfilled this role for both curriculum and assessment tools adopted and used by its districts and schools.

2. Second, the state will provide the support necessary to strengthen the human infrastructure to successfully implement and sustain school-based assessment. As discussed previously, this is an area that has been woefully overlooked in the past and one in which AI tools are poised to make great contributions.
3. Third, to a much greater extent than has been true to this point, states will focus their attention on doing what is necessary to support the appropriate interpretation and use of state assessment; that is, the state will need to devote more energy and resources to understanding and supporting instruction and the learning process. That support will include the state doing something productive with the extensive amount of data that it has at its disposal. The state, in collaboration with its stakeholders and research partners will be proactive in establishing and implementing a thorough and ongoing research agenda that includes the use of statistical modeling, data science, and AI tools to attempt to better describe the education landscape, improve predictions, and most importantly, answer key questions such as what comes next (Heck, 2024; Huff, personal communication).

For far too long, we have viewed state assessment as something distinct from the curriculum, instruction, and student learning, a one-off event, albeit one that is repeated annually. With state-supported assessment embedded in schools and instruction, the connection between assessment, curriculum, and instruction should be much clearer.

References

Betebenner, D., & DePascale, C. A. (2024). The Time Trap: Why It's Misguided to Report State Assessment Results as" Years of Learning". *National Center for the Improvement of Educational Assessment*

Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: issues and practice*, *7*(2), 5-9.

Council of Chief State School Officers. (2014). Criteria for procuring and evaluating high-quality assessments.

DePascale, C. (2015). Shall we dance: The final steps in implementing a new assessment program. Downloaded from https://charliedepascale.blog/2015/04/21/shall-we-dance-the-final-steps-in-implementing-a-new-assessment-program/.

DePascale, C. (2016). It's déjà vu all over again. Downloaded from https://charliedepascale.blog/2016/03/11/its-deja-vu-all-over-again/

Digital Wisdom Collective (2025). Human infrastructure: The hidden backbone of AI transformation. Downloaded from https://www.digitalwisdomcollective.com/blog/what-is-human-infrastructure.

Domaleski, C. (2019). Reflections on large-scale state assessment in the twenty-tens. Downloaded from https://www.nciea.org/blog/reflections-on-large-scale-state-assessment-in-the-twenty-tens/.

Heck, TJ (2024). A year of science exploration and innovation. Downloaded from https://source.cognia.org/issue-article/a-year-of-science-exploration-and-innovation/.

Hill, R.K. (2000). A success story from Kentucky. National Center for the Improvement of Educational Assessment.

Kirst, M. W., & Mazzeo, C. (1996). The Rise, Fall and Rise of State Assessment in California, 1993-1996.

Koretz, D. (1992). The Reliability of Scores from the 1992 Vermont Portfolio Assessment Program. Interim Report.

Koretz, D. M. (1996). *Perceived Effects of the Kentucky Instructional Results Information System (KIRIS)*. Distribution Services, RAND, 1700 Main Street, PO Box 2138, Santa Monica, CA 90407-2138

Koretz, D. M. (2008). *Measuring up*. Harvard University Press.

Koretz, D. M., & Barron, S. I. (1998). *The validity of gains in scores on the Kentucky Instructional Results Information System (KIRIS)*.

Moncaleano, S., & Russell, M. (2018). A historical analysis of technological advances to educational testing: A drive for efficiency and the interplay with validity. *Journal of Applied Testing Technology*, 1-19.

National Governors Association. (2010). Common core state standards. *Washington, DC*.

Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. Jossey-Bass.

Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). The Effects of Standards-Based Assessment on Classroom Practices: Results of the 1996-97 RAND Survey of Kentucky Teachers of Mathematics and Writing.

Wiggins, G. (1990). The case for authentic assessment. *Practical assessment, research, and evaluation*, *2*(1).