Testing Better: How to Improve Minnesota's Use of Assessments in Education

DONNA DRESCHER – BLUE EARTH PUBLIC SCHOOLS

KAITLIN LINDSEY – MINNEAPOLIS PUBLIC SCHOOLS

PAUL MUELLER – BROOKLYN CENTER SCHOOLS, EDUCATION MINNESOTA VICE PRESIDENT

ELIZABETH PROEPPER – PROCTOR PUBLIC SCHOOLS

DELENE SANDERS – NORTH ST. PAUL, MAPLEWOOD, OAKDALE SCHOOL DISTRICT

ERIK SIVERTSON – SAUK RAPIDS-RICE PUBLIC SCHOOLS

ANNETTE WALEN – OSSEO SCHOOL DISTRICT

LEAH ZUSTIAK – MOUNDS VIEW PUBLIC SCHOOLS

Testing Better: How to Improve Minnesota's Use of Assessments in Education

I. What is EPIC?

The Educator Policy Innovation Center, or EPIC, was founded by Education Minnesota to bring together teams of experienced educators to provide research-proven solutions to the challenges facing Minnesota schools. The teams perform a comprehensive review of academic literature on a given issue and share their own relevant experiences. After discussing the academic literature and its real-world implications for students, the educators recommend policies to meet the challenge. The coupling of sound academic research with actual classroom experience means the policies recommended by EPIC's policy recommendations are uniquely valuable because they combine the best from academia and real world practical experience.

The EPIC research teams are open to all members of Education Minnesota because practicing educators are the experts when it comes to education policy. However, the voice of the educator has often been absent in education policy discussions. CEOs, some academics, or politicians proclaim what is best for education with no grounding or experience in how their proposals affect real classrooms with actual students. As a result, our schools are hampered by disjointed, inefficient and at-times harmful state and federal policies.

Educators live every day with these policies. EPIC ensures policy makers will now have access to the best academic research and the thinking of front-line educators on the most pressing issues in education.

THE EPIC TESTING TEAM

The EPIC testing team collectively has over two centuries of teaching experience. They work with a broad range of students–from students in elementary to high school, special education, English learners, and gifted and talented. They hold a broad range of positions–classroom teachers, teaching coaches, testing specialists, and interventionists. Some of them are new to the profession while others have been teaching for decades.

Their collective experience and wisdom is reflected in their recommendations for changes in testing policy in Minnesota.

epic
EDUCATOR POLICY
INNOVATION CENTER

II. Executive Summary

There is no question that effective assessment is one of the most important parts of the educational process. Assessment–which includes everything from mid-lesson checks for understanding to state or national standardized tests–provides important information to students, educators, system leaders, and policy makers. It can be used to improve instruction, highlight gaps in outcomes between student groups, and demonstrate students' mastery of academic material.

However, our current system of assessment places too much emphasis on low-quality standardized tests administered by states and districts. These tests provide very limited useful system-level information. On balance, though, standardized tests fall short as tools for improving instruction, demonstrating the breadth or depth of student mastery, and increasing school or educator accountability. The misuse and overuse of standardized tests has the unintended effects of narrowing curriculum, focusing instruction on lower orders of thinking, and driving effective teachers away from either the profession or the students and schools who need them most.

The Educator Policy Innovation Center (EPIC) convened teachers from across Minnesota to evaluate the current role of testing in our state. Informed by research and their own decades of experience, they created several principles that can and should shape future testing policy at the federal, state, and local levels.

With respect to student learning, we suggest the following principles:

- Students deserve a robust curriculum
- Assessment should reflect what students are learning in classrooms and measure individual student growth over time
- Standardized testing should take place on an intentional schedule that does not disrupt classroom learning

With respect to teacher practice and accountability, we suggest:

- Students deserve a learning environment that provides timely and meaningful feedback
- Standardized assessments are only valuable if they provide meaningful data that educators are trained to use
- Student learning is influenced by numerous variables that include factors inside and outside of schools and classrooms
- Quality teaching is measured by numerous variables that include, but are not limited to, student performance on assessments

Denise Specht, president

"Schools should inspire and engage students while teaching them critical thinking, problem-solving and all the other skills that lead to successful lives, but that often doesn't happen due to the hyper-emphasis on standardized testing."

Education Minnesota President Denise Specht, St. Paul Pioneer Press, March 9, 2015

Based on these principles and the available research, we offer four specific recommendations:

- The state should administer fewer, better standardized tests.
- The focus on assessment should return to the classroom.
- Meaningful improvement of teaching requires peer support and a less punitive mindset.
- Communities—parents, students, teachers, and stakeholders—should be included in decisions surrounding student assessment.

This paper offers research in support both of our statement of principles and our specific recommendations. It concludes with an explanation of how these principles and recommendations better address our data needs—from student to school systems—than does the current widespread use of standardized tests at the state and district levels.

III. Statement of Principles

These principles were developed by educators from across Minnesota and reflect the realities demonstrated in the relevant research literature, interpreted through the lived experience of educators working in Minnesota schools.

STUDENT LEARNING

The ultimate goal of education is, of course, student learning. Assessment plays a key role not only in gauging the extent of that learning, but also in shaping the priorities and areas of focus for instruction. As such, it is important that our assessment practices align with our values for student learning.

STUDENTS DESERVE A ROBUST CURRICULUM

The lived experience of many students and educators is that standardized testing has created a focus on reading and math at the expense of other curricular areas. As the two earliest and most frequently assessed subjects during the modern era of standardized testing—and as the subjects carrying all or most of the weight in test-based accountability policies—tests in these two subjects have profoundly impacted the emphasis and form of classroom-level curriculum, instruction, and assessment.[1] While there is no question that reading and mathematics are foundational skills for many other disciplines, the focus they have received has reduced attention not only to other fields but also to reading- and math-specific skills and levels of learning that are difficult to assess in a standardized test.[2]

When schools are labeled as failing by accountability systems, including Minnesota's current Focus and Priority designations for schools based on the state's Multiple Measurements Rating system, they often face additional pressure to further tailor curriculum and instruction to resemble standardized tests, as well as pressure to prioritize test-aligned learning over a robust, engaging curriculum in a variety of subjects. Curriculum narrowing disproportionately impacts high-poverty schools, which are most likely to face the consequences of punishments based on test scores.[3] As a result, students in high-needs schools are the most likely to receive a narrowed education that focuses too intensely on the specific skills and lower-order question formats of mandated tests. In addition to attending classes shaped by tests, students in high-needs schools—who are more likely to be students of color—also tend to spend additional time explicitly preparing for tests.[4]

---

1 *Au, W. (2011). "Teaching under the New Taylorism: High-stakes testing and the standardization of the 21st Century curriculum." Journal of Curriculum Studies, 43(1), p. 25-45.*

2 *Berliner, D. (2011). "Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows." Cambridge Journal of Education, 41(3), p. 287-302.*

3 *Darling-Hammond, L. (2010). The flat world and education:How America's commitment to equity will determine our future. New York: Teachers College Press.*

4 *Berliner, D.C. and Nichols, S.L. (2007). Collateral damage: How high-stakes testing corrupts America's schools. Cambridge, MA: Harvard Education Press.*

This is an unacceptable inequity for our students of color, who need to be exposed to a diverse range of subjects in science, social studies, world languages, the arts, physical education, and career/technical education. Students should have the opportunity to stretch themselves into the highest levels of learning in any of these disciplines, but too often are constrained by the emphasis on shallower math and reading questions. Additionally, a well-rounded curriculum is needed to encourage students to stay in school and graduate. Curriculum narrowing in schools that disproportionately serve students of color can have the long-term effect of reducing graduation rates for students of color.

ASSESSMENT SHOULD REFLECT LEARNING

Several constraints limit the usefulness of current state-required tests for gauging the full scope and depth of student learning, suggesting that additional attention needs to be paid both to improving the quality of those tests and to guaranteeing the use of high-quality assessments throughout the school system.

The most common type of standardized assessment used by the state and by individual districts are multiple-choice tests. These tests assess lower-order thinking, not problem-solving. For example, the official test specifications for the Minnesota Comprehensive Assessments (MCAs), the primary state test in Minnesota, require only 10% of reading questions to assess "strategic thinking" and explicitly leave out the highest-level "extended thinking" questions.[5]

Assessments should also provide meaningful feedback for students and families. The current standardized tests are limited by the small range of disciplines, skills tested, a lack of depth, and long lag time before information is made useful to students and families. Finally, the overall rating of schools should reflect more learning than standardized test scores, including the availability of arts and physical education curricula, whether the school has social workers or counselors, nurses, recess or after-school programming, among other important assets. At present, the focus on the narrow set of skills and question formats included in mandatory standardized tests neglects many key aspects of student learning.

ASSESSMENT SHOULD MEASURE GROWTH

Because each student begins the school year at a different point in his or her learning, and because whole classes of students can differ from the previous years in key respects, it is essential that assessments gauge the real growth demonstrated by each individual student. This requires assessing students regularly, using that feedback both to adapt instruction and to keep students and families informed about progress, not just proficiency. While some choices can be made at the state and district levels to improve somewhat the usefulness of mandated standardized tests for measuring growth, most assessments of growth can only occur at the classroom level.[6]

---

5 *Minnesota Department of Education. (2014). "Reading test specifications for MCA-III, grades 3-8 and 10." Accessible via <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>*

6 *Anderman, E.M., Gimbert, B., O'Connell, A.A., and Riegel, L. (2014). "Approaches to academic growth assessment." British Journal of Educational Psychology.*

GRADE SPAN TESTING BEST ACHIEVES EDUCATIONAL GOALS

The focus on classroom-level assessment does not contradict the need for regular system-level assessments of educational outcomes and equity. Because of the requirements of the No Child Left Behind Act, Minnesota tests students in both reading and math in third, fourth, fifth, sixth, seventh, and eighth grades, as well as once each in high school.[7] The state also assesses science in fifth and eighth grades and once in high school. Testing in fifth grade allows for an assessment of where students are as they leave elementary school, and testing in eighth grade allows for a similar assessment at the end of middle school. This is what is known as grade span testing, an effective method of identifying significant trends and gaps in performance.

The current model of administering state tests has a dramatic impact on instruction time, but the test scores generated in third, fourth, sixth, and seventh grades do not provide enough additional information to justify their use. To be clear, statewide tests do offer information about equity gaps to districts and the state that is meaningful and necessary, but that information can be gained with a reduced level of testing. Since these tests do not offer feedback that meaningfully informs instruction, reducing their frequency to testing grade spans has the double benefit of ensuring that meta-data is available to identify trends and gaps, as well as avoid the problematic time trade-offs that annual testing creates. Furthermore, an excessive emphasis on standardized testing can encourage developmentally inappropriate state tests that consume time and resources but produce little to no meaningful information. Grade span testing could help curtail this.

TEACHER PRACTICE AND ACCOUNTABILITY

As has been thoroughly discussed in research, educators are the most significant in-school factor affecting students' performance on assessments like standardized tests (although out-of-school factors have a larger influence than in-school factors on variation in test scores).[8] These data speak to the larger issue of gauging and improving teaching quality in a way that is most helpful to student learning. Much of the current momentum in the country is to emphasize the use of tests and test-based calculations in pursuing this goal, but real improvement of teaching quality requires more depth and sophistication.

STUDENTS DESERVE A LEARNING ENVIRONMENT THAT PROVIDES TIMELY AND MEANINGFUL FEEDBACK

Research has clearly demonstrated that timely, meaningful assessment and feedback, provided regularly, is one of the most important teacher actions that improves student learning.[9] This requires the use of formative and summative assessments delivered in the classroom, as well as mid-lesson checks for understanding. (More details on these forms of

---

7 *Minnesota Department of Education. "Testing information." <http://education.state.mn.us/MDE/JustParent/ TestReq/index.html>*

8 *Nye, B., Konstantopoulos, S., and Heges, L.V. (2004). "How large are teacher effects?" Educational Evaluation and Policy Analysis, 26(3), p. 237-257.*

9 *Hattie, J., and Timperley, H. (2007). "The power of feedback." Review of Educational Research, 77(1), p. 81-112.*

assessment are included in Section IV: Focus on the Classroom.) In most cases, this feedback must be more personalized and constructive than simply letting a student know whether they selected the right answer, and in many instances, may not even be reducible to a single right answer. This goes beyond the scope of what is feasible using standardized tests and demonstrates the importance of teacher practice.

STANDARDIZED ASSESSMENTS ARE ONLY VALUABLE IF THEY PROVIDE MEANINGFUL DATA THAT EDUCATORS ARE TRAINED TO USE

Because of the importance of assessment in shaping feedback and guiding changes to instruction, educators must be well-versed in how to best collect and use data in the classroom. Time and training need to be provided for teachers to effectively analyze data and modify instruction; this is most effectively done in peer groups and ongoing professional development.[10] Especially for newer teachers whose preparation programs may not have included a sufficiently thorough assessment component and for teachers who have been constrained by heavily scripted curricula, training and retraining in effective assessment and data analysis is crucial to teachers' ability to use classroom assessments to provide students with the learning environment they deserve. Section V: Meaningful Improvement of Teaching includes further discussion of the need and options for better professional development.

STUDENT LEARNING IS INFLUENCED BY A VARIETY OF FACTORS INSIDE AND OUTSIDE OF SCHOOLS AND CLASSROOMS

Public policy related to teacher evaluation must recognize the many different circumstances that affect learning. Students from transitory populations; students who encounter tests biased against their race, culture, or economic background; and students who are required to take tests that are inappropriate for their instructional level given disabilities, familiarity with English, or other factors will see significant variations in scores that cannot be attributed solely to their teachers.[11]

However, many of the current policies under consideration for using test scores in teacher evaluation are insufficiently sensitive to the many additional factors that can affect score variation, and have therefore been deemed inappropriate for use in making high-stakes decisions about teachers.[12] Continuing to pursue a teacher accountability strategy that prioritizes these tools is not only likely to increase the share of effective teachers dismissed based on statistical flaws, but will also create an unintended disincentive for teachers to take on students from traditionally underrepresented backgrounds, students from under-resourced families and communities, students with disabilities, and students still learning English, all due to concerns about performance on high stakes tests. An improved policy must reflect these realities.

---

10 *Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., and Yoon, K.S. (2001). "What makes professional development effective? Results from a national sample of teachers." American Educational Research Journal, 38(4), p. 915-945.*

11 *Goldhaber, D.D., Brewer, D.J., and Anderson, D.J. (1999). "A three-way error components analysis of educational productivity." Education Economics, 7(3), p. 199-208.*

12 *American Statistical Association. (2014). "ASA statement on using value-added models for educational assessment," American Statistical Association. <https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf>*

QUALITY TEACHING IS MEASURED BY NUMEROUS VARIABLES THAT INCLUDE, BUT ARE NOT LIMITED TO, STUDENT PERFORMANCE ON ASSESSMENTS

The real concerns about the variables that impact student learning should not be misunderstood as a dismissal of the importance of student achievement in evaluating teachers. Minnesota's teacher development and evaluation law requires the use of multiple measures to evaluate teachers, and it provides many tools to inform areas of growth for teachers, the learning environment, and the overall school climate. Many districts have identified several different tools that can be used to gauge a teacher's impact on student learning and help teachers at all levels of experience improve their practice. This reinforces the earlier principles concerning teachers' comfort and ability using a range of assessments (including but by no means limited to conventional quizzes and tests) to track student learning.

SUMMARY

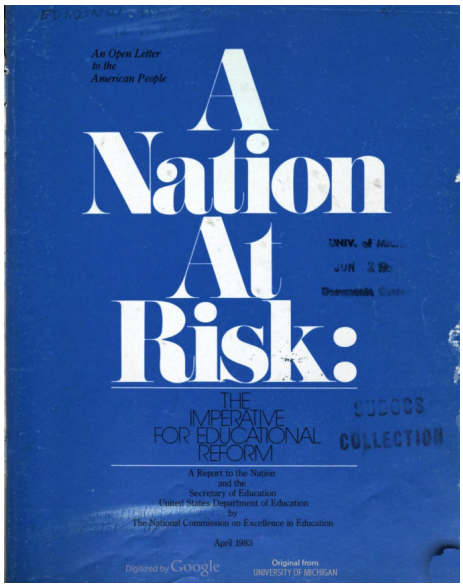
One key idea runs through all of these principles: Learning happens before and after assessment. Every part of our education system—from the commissioner of education looking at changes in statewide gaps between student populations to the third-grader trying to improve her writing skills—benefits from the well-planned use of high-quality assessments. Identifying the right assessment or assessments for each purpose is a difficult task, and we are still working to develop an approach to assessments that best puts them to use for the good of children rather than unnecessarily constraining our students, educators, and schools. The principles set forth in this section have shaped the specific recommendations to come, but they can and should also be used more broadly as a framework to evaluate assessment policies.

IV. Fewer, Better Tests

The national conversation around education is largely consumed by standardized testing. Several aspects of these tests are open to criticism, including their number and duration, their developmental appropriateness, and the motives of the small number of unaccountable corporations who sell most of them to states (along with textbooks and other preparation material to districts). However, the idea of regularly sampling student learning in key areas to identify trends and concerns has merit. The question is how best to gather this system-level data without negatively impacting the quality of education students receive.

THE CURRENT ROLE OF FEDERAL AND STATE TESTS

The present system of testing has largely evolved from two primary concerns: the United States' international competitiveness and achievement and opportunity gaps between white students and students of color.



The 1983 report, "A Nation at Risk," commissioned by the Reagan administration, framed education as a national security issue. Assessing measures of achievement in the U.S. against those of other countries, the report concluded that the state of education in the country constituted a long-term threat to national security.[13] Several of its recommendations resulted in many states adopting their own academic standards, which today provide the foundation for standardized tests in select subjects.[14] Concerns about how U.S. students measure up to other countries predate the report by decades, but "A Nation at Risk" is generally seen as having catalyzed the current approach to academic standards. Attention to international comparisons has continued, with regular attention paid to the United States' standing in assessments like the Program for International Student Assessment (PISA).

Internally, the country's focus has shifted from guaranteeing equal access to K-12 education for all students (a driving concern for much of the 20th century) to promoting educational equity in outcomes. Those concerns shaped several of the key provisions of the current incarnation of the Elementary and Secondary Education Act; that current incarnation is better

---

13 *United States. National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform : a report to the Nation and the Secretary of Education, United States Department of Education. Washington, D.C.: The Commission.*

14 *Birman, B., Bohrnstedt, G., Hannaway, J., O'Day, J., Osher, D., Phillips, G., Salinger, T. (2013). "Three decades of education reform: Are we still 'A Nation at Risk?'" American Institutes of Research. <http://www.air.org/resource/three-decades-education-reform-are-we-still-nation-risk>*

known as No Child Left Behind, or NCLB. NCLB required all states to annually test all third-through-eighth grade students in math and reading, as well as testing all high school students once.[15] The results of these tests needed to be disaggregated by student characteristics, allowing for comparisons among students of different races, economic backgrounds, disability status, and language learner status. These tests have become a focus for policy making in several respects, but their initial function was to provide data on how different groups of students fared at the system level. This is still a necessary function; the question is how best to serve it.

THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

The National Assessment of Education Progress (NAEP), administered by the federal government, is the most widely discussed national-level test in the United States. Conducted every two years in selected subjects (most prominently math and reading), the NAEP tests a representative sample of fourth, eighth, and twelfth grade students in each state.[16] It reports proficiency rates for several categories of student in each jurisdiction. NAEP scores are not connected to public policy investments or consequences, although they are sometimes invoked in policy debates.

THE MINNESOTA COMPREHENSIVE ASSESSMENTS (MCAS)

In Minnesota, the Minnesota Comprehensive Assessments (MCAs) fulfill the requirements of No Child Left Behind. They are administered annually in math and reading to all third-through eighth-graders, in reading again in tenth grade, and in math again in eleventh grade.[17] Science MCAs are conducted as grade span tests in fifth grade, eighth grade, and after a high school student has taken biology. Scores are disaggregated by student population and reported to the public. Periodically, tests in each subject are updated to reflect changes in the standards or a shift in focus. For example, past versions of the reading MCA required brief written responses by students to some questions, but the current version uses only multiple-choice questions. The level of thinking the MCAs can address is limited by the use of multiple choice questions, so higher levels of learning are rarely assessed.[18]

AN INTERNATIONAL PERSPECTIVE

It is not necessary to test as extensively as the United States does to identify system-level trends and concerns; indeed, few other developed countries do so. In fact, the National Center on Education and the Economy's 2011 report, "Standing on the Shoulders of Giants: An American Agenda for Education Reform," notes, "The idea of grade-by-grade national

---

15 *No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2003)*

16 *National Center for Education Statistics. (2014). "NAEP overview," National Center for Education Statistics website. <http://nces.ed.gov/nationsreportcard/about/>*

17 *Minnesota Department of Education. "Testing information." <http://education.state.mn.us/MDE/JustParent/ TestReq/index.html>*

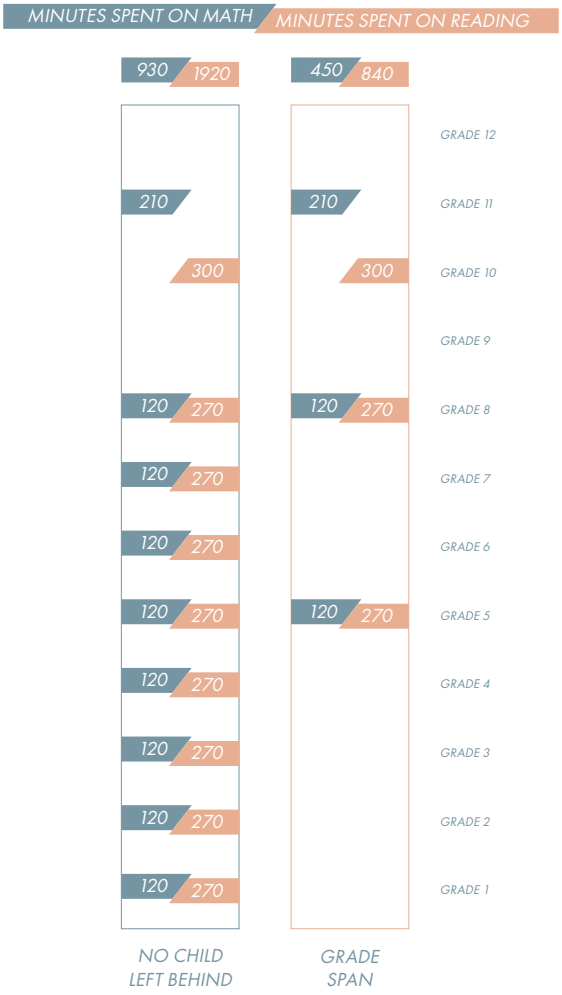
18 *Minnesota Department of Education. (2014). "Reading test specifications for MCA-III, grades 3-8 and 10." Accessible via <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>*

testing has no takers in the top-performing countries."[19]

This is not to say that high-scoring countries do not conduct system-level testing. For example, Amanda Ripley's The Smartest Kids in the World, and How They Got That Way notes that, for system-level assessment, Finland tests a representative sample of students in a small number of grades, in similar fashion to the NAEP.[20] The NAEP appears to be in line with healthy practice elsewhere, and is a tool for federal policy-makers and those studying gaps between student populations across states.

TIME SPENT TESTING

NO CHILD LEFT BEHIND V. GRADE SPAN TESTING



Times estimates from 2014-2015 MCA Procedures Manual, St. Paul Public Schools

19 *Tucker, M.S. (2011). "Standing on the shoulders of giants: An American agenda for education reform," National Center on Education and the Economy, p. 8. <http://www.ncee.org/wp-content/uploads/2011/05/Standing-on-the-Shoulders-of-Giants-An-American-Agenda-for-Education-Reform.pdf>*

20 *Ripley, A. (2013). The Smartest Kids in the World, and How They Got That Way. Simon & Schuster: New York, NY.*

RECOMMENDATIONS

As suggested by the title of this section, improving the system-level use of standardized tests will require fewer, better tests. Again, it is important to collect this high-level data to track our progress on achieving education equity for all students. The question is how best to do so. As the NAEP appears sufficient for this goal at the national level (producing high-quality data without imposing an unnecessary burden on students or narrowing curriculum), the focus of these recommendations will be on the MCAs in Minnesota. The MCAs are particularly important for identifying districts and schools for attention and support (but not shame or punishment), since the representative sampling techniques used by the NAEP do not allow for that level of detail.

TRANSITION THE MCAS TO GRADE SPAN TESTING IN FIFTH GRADE, EIGHTH GRADE, AND HIGH SCHOOL

Minnesota already uses grade span testing for its science MCAs. The information gained from these tests allows policy makers and school leaders to identify gaps between student populations at the state, district, and school levels. Moving reading and math testing to a similar approach would reduce the number of hours students spend testing, the time that computer labs are unavailable because they are dedicated to testing and potentially reduce attention to test preparation without sacrificing attention to the very real gaps between students based on race, family income, disability status, and language learner status.

Since the MCAs do not provide meaningful information to students, families, or educators that could not be generated through deeper, more nuanced classroom assessment and feedback, moving from annual testing to grade span testing does not sacrifice much, if any, useful information. (Section IV: Focus on the Classroom describes many other sources of assessment data that offer much greater detail and direction to families than do the MCAs.) Grade span testing will still result in the vital, annual attention to proficiency gaps as measured by tests, but will free up time and resources that can be put to better use.

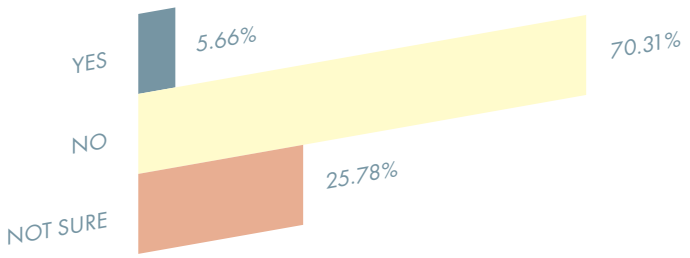
REDUCE THE AMOUNT OF TIME SPENT TAKING STATE TESTS

At present, students in most Minnesota schools are likely losing weeks to a month or more of instruction time either taking tests or engaged in explicit test preparation. (While full statewide data are unavailable and districts can vary in the amount of time spent on test preparation, these estimates are based on national research, published testing schedules, the experiences of educators across Minnesota, and feedback reported to the Minnesota Department of Education.)[21, 22] The legislature took an important first step to limiting testing time in 2015 by capping hours spent on district standardized tests at 10 hours. The Commissioner of Education and Governor Dayton should be lauded for their leadership on this issue. Further efforts to better enforce this limit, as well as reduce the threat of sanctions for schools and teachers would result in a more meaningful reduction of time spent on standardized testing in the long run.[23]

IMPROVE THE QUALITY OF THE MCAS

In addition to creating more time for learning, a move to fewer tests allows more funds to be invested in creating better tests. As already discussed, the levels of learning assessed by the MCAs exclude the highest levels of learning, and the scope of testing is limited to those skills that can be assessed in a multiple-choice format. If some of the funds currently spent on the many tests administered in third, fourth, sixth, and seventh grades were channeled into creating richer, more nuanced assessments using a wider variety of question type and assessing higher levels of learning in fifth grade, eighth grade, and high school, the overall quality of data produced by statewide testing would increase.

DO YOU TRUST THE SCORES FROM THIS YEARS' MCA TESTS WILL ACCURATELY REFLECT STUDENT LEARNING.



Education Minnesota surveyed its members in April 2015 about the implications of the extensive technical problems with the administration of the Minnesota Comprehensive Assessments that year. An overwhelming number of educators said that they wouldn't trust the results of the test.

---

21 *Nelson, H. (2013) "Testing more, teaching less: What America's obsession with student costing costs in money and lost instructional time," American Federation of Teachers.*

22 *Minnesota Department of Education. (2012). "Feedback: Testing in districts," Minnesota Department of Education.*

23 *Cassellius, B. February 2, 2015 letter to legislators. Digital copy viewable at <https://s3.amazonaws.com/s3.documentcloud.org/documents/1683988/testing-reduction-advisory-group-cover-letter.pdf>*

Such a reinvestment would also offer an opportunity for the state to cultivate a public source of tests rather than paying one of the few large testing corporations that create, deliver, and score tests (as well as sell textbooks and test preparation materials). Currently, Minnesota has a three-year, $38 million contract with Pearson Education, a billion-dollar testing corporation, to administer MCAs. In 2015, glitches, problems logging in and crashes caused Education Commissioner Brenda Cassellius to suspend all state testing until the problems were addressed. Despite the suspension & legislative attempts to do so, Minnesota has very limited ability to hold Pearson accountable for the errors.

Minnesota's institutions of higher learning and research organizations contain experts on assessment, and Minnesota's community leaders and educators can speak to what matters most for student learning. Using these resources in a reinvention of statewide testing that emphasizes quality over quantity would benefit everyone involved more than the use of comparatively shallow tests purchased from testing corporations.

ALIGNMENT WITH PRINCIPLES

As with all of the recommendations in this paper, these are based on the principles articulated by educators.

STUDENTS DESERVE A ROBUST CURRICULUM

To the extent that testing informs choices about curriculum and instruction, creating higher-quality tests will encourage school and district leaders making choices about curriculum purchases to focus on higher orders of thinking and more varied approaches to assessment. The current emphasis on narrow tests with significant stakes attached can and should be replaced by a system that values deeper learning while still informing policy-makers and the public about Minnesota's progress towards educational equity for all student populations.

ASSESSMENT SHOULD REFLECT LEARNING

Higher quality tests will also better reflect students' capacity for problem solving and offer more meaningful feedback to students and families. Trading the frequency of testing for quality will also encourage the broader view of schools and districts that considers availability of other forms of learning, such as arts and physical education programs.

STANDARDIZED TESTING SHOULD TAKE PLACE ON AN INTENTIONAL SCHEDULE THAT DOES NOT DISRUPT CLASSROOM LEARNING

Once again, the importance of system-level testing to sustaining attention on the many gaps that hurt Minnesota's children of color and Minnesota's children from under-resourced families and communities cannot be overstated. However, the shallow annual testing of the MCAs is neither necessary nor best-suited to that task. The move to grade span testing provides similar information, but allows for better quality and reduces the substantial negative impact on instruction time endemic to the current system.

SUMMARY

The foundational purpose of system-level testing in Minnesota is to make public the trends and concerns facing the state, its individual districts, and their individual schools. This is a necessary purpose, especially as we focus on correcting the inequities that pervade our society and our school system. However, the use of fewer, better tests can improve the quality of information produced by that system-level testing and reduce some of the unintended consequences of the current system on curriculum and instructional time. It is also important to recognize that system-level testing is just one of the levels of assessment, and one of the least impactful for improving student learning. For that, we must turn to the classroom.

V. Focus on the Classroom

For as much attention placed on state-mandated testing, most educational assessment happens at the classroom level. Placing more attention on the quality of in-school assessment is critical to improving student learning.

THE IMPORTANCE OF IN-SCHOOL ASSESSMENT

While large-scale standardized tests like the MCAs can provide important information about system-level outcomes, they provide little to no practical information for improving instruction.[24] Even in cases where particular skills can be identified as stumbling blocks for specific student populations at individual schools, the tests provide no specific information about how to better teach those skills. That information can only come from regular assessment of students in the classroom and from careful analysis of that assessment data to improve instruction, target remediation, and identify successful teaching techniques.

THE GOALS AND METHODS OF CLASSROOM ASSESSMENT

Classroom assessment can take several forms. Most immediately, a teacher can check for students' understanding midway through a lesson.[25] This can be done using any of a number of techniques such as collecting student responses through the use of personal whiteboards, color-coded chips, simple hand-raising, or specialized technology, to name a few. Teachers can respond to this data in several ways, including re-teaching the whole class in the case of widespread misunderstanding or grouping students by mastery later in the lesson (with an opportunity to reteach, review, or enrich the particular knowledge or skills in question).

Teachers can also collect data at the end of each lesson, for example, in the form of exit slips, to create a larger and more durable picture of student understanding.[26] These data can shape plans for future lessons, identify particular students in need of individual attention outside of class, and be shared with peers as part of professional conversations about how the lesson worked.

Formative assessments can assess student mastery of material from multiple lessons and provide an opportunity for feedback from the teacher before the end of a larger unit or topic.[27] Again, many types of formative assessment are available, from the familiar quiz format to shorter writing projects to lab experiments and many other hands-on learning experiences that still produce a product or performance that the teacher can assess. Several factors contribute to effective formative assessment, including but not limited to: its alignment with what was actually taught, students' understanding of how they are being assessed, the

---

24 *Wiggins, G. (1990). "The case for authentic assessment." ERIC Digests. ERIC Clearinghouse on Tests, Measurement, and Evaluation; American Institutes for Research.*

25 *Rosenshine, B.V. (1986). "Synthesis of research on explicit teaching." Educational Leadership, 43, p. 60-69.*

26 *Black, P., Harrison, C., Marshall, B., Wiliam, D., & Lee, C. (2003). Assessment for Learning: Putting it into Practice. Buckingham: Open University Press.*

27 *Ibid*

speed with which the teacher gives feedback to students, and the quality of that feedback.[28]

Summative assessments cap off a larger unit or topic of study (and sometimes an entire semester or year).[29] The goal of summative assessments is to identify enduring student mastery of previously taught skills. Unit tests, papers, and individual projects are just some of the methods available for summative assessment, and the information gained from those assessments can help teachers adapt upcoming instruction, follow up with individual students, and refine their techniques for the future. It should also be noted here that the MCAs should not be used as summative assessments because they test only a selection of standards, fail to assess the highest levels of learning, and do not provide sufficient scaffolding of feedback to aid educators or students.

All of these classroom assessments provide more immediate, thorough, relevant, and often higher-order thinking data than standardized tests. Classroom assessments are the real backbone of education, and should be emphasized more strongly.

THE RESEARCH ON CLASSROOM LEVEL ASSESSMENT AND FEEDBACK

Research on teacher actions indicates that providing relatively quick and very constructive feedback is one of the most important steps teachers can take to support student learning.[30] Contrary to the popular conception of teaching that emphasizes lecturing and teacher-centered content delivery, a significant share of student learning comes from using their new knowledge and skills, making mistakes, and receiving helpful direction to improvement. This may explain part of why the speed and quality of feedback is so important for student learning.

Classroom assessment allows teachers to create nuanced evaluations that scaffold the skills students need to use so that they can better identify where individual students have misunderstood or are otherwise struggling with the content they have taught.[31] This characteristic of classroom assessments partially explains why they are often more useful for teachers than shallower standardized tests. Additionally, classroom assessments that deviate from the multiple choice format allow teachers to better gauge students' higher-order learning. This is why the guidelines for the MCAs specifically state that higher orders of thinking must be assessed in the classroom rather than through the state test.[32]

As will be discussed in more detail in Section V: Meaningful Improvement of Teaching, regularly assessing students in the classroom is also a key component to effective professional development for individual teachers.

---

28 *Hattie, J., and Timperley, H. (2007). "The power of feedback." Review of Educational Research, 77(1), p. 81-112.*

29 *Harlen, W., and James, M. (1997). "Assessment and learning: Differences and relationships between formative and summative assessment." Assessment in Education: Principles, Policy & Practice, 4(3), p. 365-379.*

30 *Hattie, J., and Timperley, H. (2007). "The power of feedback." Review of Educational Research, 77(1), p. 81-112.*

31 *Moon, T.R. (2005). "The role of assessment in differentiation." Theory Into Practice, 44(3), p. 226-233.*

32 *Minnesota Department of Education. (2014). "Reading test specifications for MCA-III, grades 3-8 and 10." Accessible via <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>*

THE ROLE OF DISTRICT-SELECTED STANDARDIZED TESTS

Recent years have seen many districts in Minnesota purchasing additional standardized tests for local use.[33] One particularly popular example is the NWEA MAP test, but most of these tests share the same features: They are delivered online, adapt their difficulty based on students' answers, provide quick data for teachers and administrators to use, and are typically administered two or three times a year to gauge growth. Like the state-mandated standardized tests they are meant to emulate, most district-selected standardized tests rely heavily or exclusively on multiple choice questions.

Educators and administrators across Minnesota have reported varying degrees of satisfaction with these tests. Given the heavy policy emphasis on state tests, these district-selected tests are often seen as a way to identify ahead of time which students need support in specific areas of math and reading. (Although other subjects are also often available, many schools and districts choose to focus on math and reading.) Schools and districts have engaged productively with educators in many cases to target interventions for students. In other cases, the costs of the tests, not just in money, but also in lost instructional time and reduced access to the computers needed to administer the tests, are seen as too high for the value of the information gained.

In many ways, district-mandated standardized tests address a symptom of high-stakes state testing, which is the lack of timely and useful data, but they do not solve the underlying issues with state testing.

RECOMMENDATIONS

Ultimately, the use and improvement of in-school assessments is the responsibility of educators and other staff in individual schools and districts. Still, policy-makers and advocates can do more to emphasize the importance of in-school assessment and encourage high-quality assessment at the classroom and school levels, while de-emphasizing flawed standardized tests.

SUPPORT TEACHERS IN CREATING, USING, INTERPRETING, AND IMPROVING A RANGE OF ASSESSMENT TOOLS

The bulk of in-school assessment is conducted by teachers. In some cases, packaged curricula provide some or all of the assessments a teacher will use. In other cases, teachers have total control (and total responsibility) for determining, designing, using, and refining the assessments in their classrooms. Regardless of the degree of teacher power over assessment, more support must be given to teachers to maintain and develop their ability to create, use, interpret, and improve the full range of assessment tools discussed in this section.

Several organizations and roles can offer aspects of this support. Teacher preparation programs should include serious study of assessment (including discipline-specific assessment) and data analysis to prepare teachers to be competent in these areas regardless of the

---

33 Minnesota Department of Education. (2012). "Feedback: Testing in districts," Minnesota Department of Education.

individual school environment they enter. School and district leaders can use professional development time and resources to focus on high-quality assessment practices and involve classroom-level educators in the process of selecting, developing, and presenting professional development options for this purpose. Schools and districts that use instructional coaches can provide training as necessary to ensure that coaches are prepared to support teachers in using assessment constructively.

Additionally, observation and evaluation systems can recognize the importance of assessment as a teacher skill and provide constructive feedback and, as appropriate, ongoing support in improving assessment practices. State and local policy-makers can invest more in professional development as well as publicly recognize the limitations of state-mandated tests with respect to improving either instruction or learning.

INVOLVE EDUCATORS IN DECISIONS ABOUT DISTRICT-LEVEL STANDARDIZED TESTING

With respect to district-selected tests like the NWEA MAP, educators must be involved in decisions about whether and how to use these tests, as well as discussions about what classroom-based alternatives might exist. This likely means that some districts choose to continue using these tests, while others will return the responsibility for assessing students' math and reading mastery and growth to the classroom level. Ultimately, though, educators must play a significant role in this decision making process.

ALIGNMENT WITH PRINCIPLES

As with all of the recommendations in this paper, these are based on the principles articulated by educators.

STUDENTS DESERVE A ROBUST CURRICULUM

Classroom assessment allows for analysis of a greater breadth and depth of learning than does state-mandated standardized testing. Policy and practice at several levels of our education system can encourage meaningful improvement for students' learning by placing greater emphasis on supporting educators in developing their assessment skills. The only way we can trust that students are receiving the robust curriculum they deserve is by knowing that educators all across the state have the ability to use classroom assessments to demonstrate and improve student mastery of many skills beyond what the MCAs can sample.

ASSESSMENT SHOULD REFLECT LEARNING

It is educators' responsibility to make sure that the assessments they are using in their schools and classrooms truly reflect student learning. Doing so allows them to provide much better feedback to both students and families about students' mastery of many skills and ability to use higher orders of thinking.

ASSESSMENT SHOULD MEASURE GROWTH

District-selected tests can be a starting point for gauging student growth over the course of the year. In most disciplines, however, that assessment can only happen at the classroom

level. (Even in disciplines included in the district tests, educators will often need to supplement what the test measures; for example, English teachers need to use other methods to assess most writing and research skills.) As a result, the primary place this principle can be satisfied is in the classroom.

STUDENTS DESERVE A LEARNING ENVIRONMENT THAT PROVIDES TIMELY AND MEANINGFUL FEEDBACK

Research has demonstrated that timely, meaningful feedback is critical to student learning. That must go beyond telling a student how many right answers they got for tested subjects and skills. For students to have the learning environment they deserve, their teachers must be able to assess and provide feedback regularly and with high quality.

STANDARDIZED ASSESSMENTS ARE ONLY VALUABLE IF THEY PROVIDE MEANINGFUL DATA THAT EDUCATORS ARE TRAINED TO USE

Too often, educators are overwhelmed with data or insufficiently prepared to collect and analyze data independently. Appropriate training in data analysis must be part of the increased focus on effective use of data, not just from state and district tests but from classroom assessments as well.

SUMMARY

Perhaps the most important change we can encourage in the public discussion of testing is a shift from system testing, which does provide necessary data but cannot do much to improve quality, to in-school assessment. On the simplest level, it must be made clear that teachers conduct assessments frequently over the course of each school day, week, and year, and that state tests are some of the least helpful for guiding improvements in student learning. We must be able to trust educators' use of these classroom assessments, which means supporting them in developing and maintaining the many specific skills necessary to assess students effectively and respond to assessment results. Creating and preserving this trust means paying attention to classroom assessment and identifying beneficial tools for helping teachers in this area.

VI. Meaningful Improvement of Teaching

One of the clear implications of the analysis in Section IV is that Minnesota needs to continue investigating the best tools for improving teaching. The subset of skills for high-quality assessment and feedback is just one component of the larger set of skills needed for effective teacher practice. Finding constructive ways to encourage improvement of those skills, both for individual teachers and for the teacher corps as a whole, continues to be a source of questions and occasional frustration, but also of hope.

CURRENT FEDERAL AND STATE APPROACHES TO IMPROVING TEACHING

At present, much of the high-level attention to improving teaching is focused on the use of standardized tests and other sources of data to score teacher performance. While the use of student data does have a role to play in discussions of teaching quality, many of the tools receiving the most attention right now are focused on incentives and punishments, not on supporting improvement. Additionally, much of the focus on test-based systems emphasizes proficiency rates, either in absolute terms or comparing a teacher's class from one year to their class the next year, with insufficient attention to student growth.

NEW TEACHER EVALUATION SYSTEMS

Minnesota is one of many states to adopt a statewide teacher evaluation law. These laws were encouraged by the federal Department of Education as it determined states' eligibility for waivers from the harsher restrictions of No Child Left Behind.[34] One of the highest-profile components of the law is the requirement that student data (generally seen as meaning

34 *U.S. Department of Education. (2012) "ESEA flexibility policy document." Accessible at <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>*

standardized test scores) be incorporated into the evaluation of teachers.[35] Districts have the flexibility to define local evaluation systems, but those local systems must still meet certain criteria established by the state, including the use of student data as part of teacher evaluation.

The use of smaller-scale student data is more helpful to teachers because it can be combined with outside observation by an administrator or coach. For example, when an observer can analyze lesson-specific student results with a teacher following an observation, the observer can offer the teacher constructive feedback and support the teacher in developing data analysis skills (as suggested in Section IV: Focus on the Classroom). The use of larger summative data to track both proficiency and growth can help a teacher, with or without the support of a coach or qualified administrator; draw additional conclusions about their practice and how to improve it in the future. Furthermore, the discussions of what constitutes appropriate data in fields beyond those covered in standardized tests have produced many constructive professional conversations among educators, as well as among teachers and other school and district personnel. Clearly, incorporation of student data into teacher evaluation can be done constructively with an eye toward helping current teachers improve. (It is also important to note the distinction between observations by coaches, which are solely growth focused, with those by administrators that include an aspect of performance review as well as attention to improvement.)

THE USE OF TEST SCORES AND "VALUE-ADDED MEASUREMENT" IN EVALUATING TEACHERS

By contrast, simply condensing student results into part of an evaluation does not offer a teacher direction on how to improve. Instead, the function of the current system is to more easily rank teachers based on their evaluation scores. While some see this only as a tool for differentiating teacher quality, others seek to use these scores to guide human resources decisions such as layoff order (or in more extreme cases, pay and retention). It is worth noting that research on the effects on linking bonus pay to test scores has suggested that the practice does not have a meaningful impact on teaching quality.[36] Additionally, there is ample research that ranking employees is counterproductive, hurts collaboration, and reduces overall employee performance. Microsoft and other large corporations have moved away from such rankings because of their detrimental impact on employees.

Other jurisdictions have embraced the use of so-called "value-added measurement" (VAM) tools when evaluating teachers. These calculations use test scores in an attempt to estimate an individual teacher's effect on variations in student performance. At present, these tools are unpredictable; slight changes in the way they adjust for student demographics, prior year's performance, and other factors can produce significant shifts in teachers' evaluation scores. Publication of these scores, especially given their questionable accuracy, has been criticized

---

35 *Post, T. (2013). "New evaluations put Minnesota's teachers to the test." The Hechinger Report. <http://hechingerreport.org/new-evaluations-put-minnesota-teachers-to-the-test/>*

36 *Marsh, J.A., Springer, M.G., McCaffrey, D.F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., Peng, A. (2011). "A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report," The Rand Corporation. <http://www.rand.org/pubs/monographs/MG1114.html>*

by some experts involved in the development of VAM systems.[37]

Furthermore, the American Statistical Association remains skeptical about the use of current VAM methods in accountability systems. In the April 8, 2014, "ASA Statement on Using Value-Added Models for Educational Assessment," the ASA included the following points:

- Estimates from VAMs should always be accompanied by measures of precision and a discussion of the assumptions and possible limitations of the model. These limitations are particularly relevant if VAMs are used for high-stakes purposes.
- VAMs are generally based on standardized test scores, and do not directly measure potential teacher contributions toward other student outcomes.
- VAMs typically measure correlation, not causation: Effects, positive or negative, attributed to a teacher may actually be caused by other factors that are not captured in the model.
- VAM scores and rankings can change substantially when a different model or test is used, and a thorough analysis should be undertaken to evaluate the sensitivity of estimates to different models.
- VAMs should be viewed within the context of quality improvement, which distinguishes aspects of quality that can be attributed to the system from those that can be attributed to individual teachers, teacher preparation programs, or schools. Most VAM studies find that teachers account for about 1% to 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions. Ranking teachers by their VAM scores can have unintended consequences that reduce quality. (p. 1-2)[38]

For the time being, VAM tools and other test-based mechanisms appear questionable at best as mechanisms for governing human resources practices. Additionally, as discussed earlier, they do not offer any meaningful path to improvement for current teachers.

AN INTERNATIONAL PERSPECTIVE

As with the country's use of annual standardized testing, the United States' experimentation with test-based teacher evaluation deviates from the practices of countries with better performance on international assessments (which, it should be noted, tend to require much deeper learning than most standardized tests used in this country).[39]

Whereas much of the founding energy for building more student data into teacher evaluations and using those evaluations for layoff, pay, and retention decisions comes from a desire to remove certain teachers and place new ones, other countries' educational cultures put more emphasis on ongoing improvement and the proliferation of useful teaching practices. One particularly dramatic example of this is Japan, where effective teaching

---

37 *Butrymowicz, S. and Garland, S. (2012). "How New York City's value-added model compares to what other districts, states are doing," The Hechinger Report. <http://hechingerreport.org/how-new-york-citys-value-added-model-compares-to-what-other-districts-states-are-doing/>*

38 *American Statistical Association. (2014). "ASA statement on using value-added models for educational assessment," American Statistical Association. <https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf>*

39 *Tucker, M.S. (2011). "Standing on the shoulders of giants: An American agenda for education reform," National Center on Education and the Economy, p. 8. <http://www.ncee.org/wp-content/uploads/2011/05/Standing-on-the-Shoulders-of-Giants-An-American-Agenda-for-Education-Reform.pdf>*

techniques originally developed in the United States have spread even as they go largely undiscovered here. In Japan, teachers spend somewhat less time in front of students so they can observe their peers and collaborate more actively in developing, implementing, and refining teaching techniques and skills.[40] This fosters a system of continuous improvement and social accountability to one's peers, and it is routinely informed by student data.

RECOMMENDATIONS

As Minnesota's teacher evaluation law and local variants grow more established, state and local policy must pay more attention to encouraging meaningful improvement of teaching quality. Especially as Minnesota faces an upcoming demographic shift that may produce widespread teacher shortages,[41] improvement of teaching quality must be seen primarily as a matter of concretely improving teachers' skills, with changes to layoff and due process practices an even less useful approach than is currently the case.

IMPROVE THE QUALITY AND TARGETING OF PROFESSIONAL DEVELOPMENT

The term "professional development" can encompass a wide range of approaches. Too often, it takes the form of "sit-and-get" lecturing from outside experts, with minimal individualization or follow-up. Research suggests, however, that sustained attention to specifically targeted areas for professional development, especially in collaboration with peers, can produce real, lasting changes in classroom practice.[42] State, district, and school leaders should also work with educators to identify those who are already local or regional experts in key skills, using their expertise to support other educators identified as needing targeted support.

EXPAND OPTIONS FOR PEER SUPPORT AS PROFESSIONAL DEVELOPMENT

While many schools have adopted some version of peer-based Professional Learning Communities (PLCs), often PLC practice falls far short of the original design.[43] Whatever the form, the ongoing use of peers for professional development is an important step in creating a professional culture of shared investment in improvement. Key aspects of this include time to observe other teachers, both to provide feedback and to learn, as well as support and training in analyzing student data in connection with teacher practice (as discussed in Section IV: Focus on the Classroom).

ENCOURAGE THE DEVELOPMENT OF HIGH-QUALITY, PEER-BASED APPROACHES TO ACCOUNTABILITY

The first priority of teacher improvement policy must be helping teachers improve in areas where they are struggling. Those who continue to struggle despite support, however, may

---

40 *Green, E. (2014). Building a Better Teacher: How Teaching Works (and How to Teach It to Everyone). W.W. Norton & Co: New York, NY.*

41 *Minnesota State Demographic Center. (2013). "In the shadow of the Boomers: Minnesota's labor force outlook." <http://mn.gov/admin/images/in-the-shadow-of-the-boomers-labor-force-outlook-msdc-dec2013.pdf>*

42 *Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., and Yoon, K.S. (2001). "What makes professional development effective? Results from a national sample of teachers." American Educational Research Journal, 38(4), p. 915-945.*

43 *Talbert, J.E. (2010). "Professional learning communities at the crossroads: How systems hinder or engender change." Second International Handbook of Educational Change; Springer International Handbooks of Education, 23, p. 555-571.*

ultimately need to leave the classroom for the sake of student learning. Districts elsewhere in the country, such as Montgomery County, Maryland, as well as here in Minnesota, such as Saint Paul, have effectively used Peer Assistance and Review (PAR) systems to help teachers at all levels of experience and skill improve, as well as to guide consistent under-performers out of the classroom.[44] This is one option for increasing the role of peer-based accountability, and it is a more informed and professional approach than reliance on test-based algorithms.

USE SOURCES OF INFORMATION ABOUT STUDENT PERFORMANCE OTHER THAN STANDARDIZED TEST SCORES AND VAM CALCULATIONS

Because of the limited nature of test scores and VAM calculations as indicators of an individual teacher's skills, those sources of data should not be emphasized in the teacher evaluation process. Teachers outside of tested subjects and grade levels have invested significant effort in identifying useful sources of student data to support Minnesota's evaluation process; teachers who are currently evaluated based primarily on test scores and VAM calculations should be allowed and encouraged to use similarly useful sources.

ALIGNMENT WITH PRINCIPLES

As with all of the recommendations in this paper, these are based on the principles articulated by educators.

STANDARDIZED ASSESSMENTS ARE ONLY VALUABLE IF THEY PROVIDE MEANINGFUL DATA THAT EDUCATORS ARE TRAINED TO USE

A greater emphasis on using student data to improve specific teaching skills requires teachers to have appropriate training in data analysis. This is important whether they are working with administrators, coaches, or peers. These recommendations would include such support in the professional development process.

STUDENT LEARNING IS INFLUENCED BY NUMEROUS VARIABLES INSIDE AND OUTSIDE SCHOOLS AND CLASSROOMS

Excessive reliance on test-based tools for evaluation and accountability is likely to have perverse consequences on those who teach or are considering teaching students from backgrounds that tend to have a negative impact on test scores. This is especially true for student populations, such as students who are English learners and those receiving special education services, that are frequently required to take tests that are inappropriate for their instructional level. Thus, these recommendations reflect the fact that more useful student data exists, and that it can be used much more constructively in meaningfully improving teaching.

QUALITY TEACHING IS MEASURED BY NUMEROUS VARIABLES THAT INCLUDE, BUT ARE NOT LIMITED TO, STUDENT PERFORMANCE ON ASSESSMENTS

Minnesota's evaluation system uses multiple measurements to evaluate teachers, and the recommendations here support such an approach. The process of evaluating and developing

---

44 *Papay, J.P. and Johnson, S.M. (2012). "Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs." Educational Policy, 26(5), p. 696-729.*

teachers uses an abundance of information to inform areas of growth for teachers, the learning environment, and school culture, and these recommendations respect and sustain that fact.

SUMMARY

While much of the national discussion about teacher evaluation revolves around incentives and punishments, Minnesota is well-positioned to make significant strides in improving teaching while respecting both statistical reality and the professionalism of its thousands of educators. Focusing on skill improvement, empowering teachers to support and hold each other accountable as peers, and evaluating teachers based on information more meaningful than can be gleaned from test scores are all useful steps that policy-makers and public leaders can encourage.

VI. Including the Community

Up until this point, the recommendations made in this paper have focused on the traditional members of the school system: educators, district leaders, state policy-makers, etc. However, assessment is one of many areas where greater inclusion of school's communities can invite real change and foster stronger connections among educators, families, and community leaders. These recommendations, which focus on including the community in decisions about assessment, will work best in the context of the full-service community schools or Promise Neighborhood framework, but can be applied productively by any school or district.[45]

THE BENEFITS OF COMMUNITY INCLUSION

While some options already exist for community members to participate in their public schools, voting for school board, joining the PTA, or sitting on a site council (if it exists), to name a few, it is rare for community members to engage frequently and constructively with educators around issues of assessment. However, those schools that have pursued such engagement have often seen positive results.[46]

---

45 *Blank, M. (2009). "Community schools and Promise Neighborhoods." Coalition for Community Schools at the Institute for Educational Leadership. <http://www.communityschools.org/assets/1/AssetManager/ PromiseNeighborhoods_ CommunitySchools.pdf>*

46 *Langberg, J., Qureshi, T., and Deas, E. (2013). "Community-based accountability: Best practices for school officials." Poverty & Race Research Action Council, 22(2), p. 9-12.*

DETERMINING GOALS AND TOOLS

One of the first steps of community inclusion concerning assessment is for educators and community members to collaborate on determining what goals (both non-academic and academic) are most important for the school to pursue and what tools are best suited to gauging progress toward those goals.[47]

For example, community members and educators might agree that, in addition to the state's academic standards, they would like to see students improve non-cognitive skills like persistence and attentiveness (skills that are useful for students in college and in the workplace). They might then agree on one or more tools, rubrics, qualitative observations, easily gathered statistics, etc. that can indicate how students are doing in those areas as well as in their academics.

The process of identifying goals and tools must be collaborative. Neither the school nor the community should simply dictate to the other what will be prioritized. Indeed, one of the key mindsets for this process is that educators and the community are not groups in opposition or isolation, but rather are partners with different information and expertise that can help inform the process.

MAKING ADJUSTMENTS BASED ON RESULTS

As educators and community members use the agreed-upon tools to check progress toward their goals, it may be necessary at times to adjust their plans or course of action. As with a teacher using classroom assessment data to modify upcoming lessons and revise plans for next year, educators and community members should be willing to make changes when necessary.

INCREASING SHARED OWNERSHIP AND INVESTMENT

The result of this process is a greater sense of partnership between educators and community members. Having worked together to set and pursue goals, adults in all parts of the process will feel a greater sense of accountability to one another for seeing student results. As a result, students are more likely to get the help and support they need both in and out of school.[48]

RESEARCH ON COMMUNITY INCLUSION

The research on community inclusion suggests that, when done properly and in good faith, it can lead to significant gains for students in both academic and non-academic areas. Individual schools in Minnesota that have pursued this approach have seen progress, and this process can also be a key part of a larger community schools or Promise Neighborhood framework that brings additional services and resources into the picture to get even more dramatic results.

---

47 *Blank, M.J., Melaville, A., and Shah, B.P. (2003). "Making the difference: Research and practice in community schools." Coalition for Community Schools at the Institute for Educational Leadership.*

48 *Ibid*

RECOMMENDATIONS

This is necessarily a more local practice, and any mandate to use it is likely to fail. Still, some concrete recommendations are possible. It should also be noted here that schools bear a significant share of the responsibility for prompting and sustaining community involvement, recognizing the many barriers to participation that families can face, including but not limited to: the need to work multiple jobs or jobs with inconsistent or unconventional hours; transportation challenges; and negative or harmful past interactions with the school system. These are not excuses to ignore community engagement, but rather realities that push schools to pursue alternative methods of engagement such as meeting outside the typical school or business day and in neutral spaces (i.e. neither the school itself nor families' homes, especially for adults who have reason not to trust the school system).

SUPPORT THE CREATION OF SCHOOL ASSESSMENT COMMITTEES

These committees, which should be created and implemented in local school districts, should be composed of parents, educators, and stakeholders. Their main charge, as outlined above, would be to identify district assessment goals, help guide assessment decisions, and revise assessment policies at the local level.

CELEBRATE LEARNING BEYOND STANDARDIZED TESTS

Getting the most out of community engagement means a willingness to look beyond standardized tests. In fact, celebrating a wide range of learning encourages more creative thinking from the collaborators and is thus likely to produce wider support for students.

ALIGNMENT WITH PRINCIPLES

As with all of the recommendations in this paper, these are based on the principles articulated by educators.

STUDENTS DESERVE A ROBUST CURRICULUM

Engaging with the community can be a significant source of validation for a curriculum that extends well beyond what is emphasized in standardized tests. It is also a way to ensure that the community as a whole is better informed and more strongly committed to the full breadth and depth of learning schools can offer.

ASSESSMENT SHOULD REFLECT LEARNING

Engaging with the community to identify goals and tools increases the opportunities to align local assessments with the many different types of learning community members and educators choose to pursue.

STUDENT LEARNING IS INFLUENCED BY NUMEROUS VARIABLES INSIDE AND OUTSIDE OF SCHOOLS AND CLASSROOMS

Ongoing and constructive engagement with community members can give educators more insight into the specific factors that affect student learning, and can spur actions on the part of schools, families, or community organizations to respond to those factors without using them as an excuse for educational inequity.

SUMMARY

At the core of our public school system is the idea that schools should reflect the needs and desires of the public. One way to encourage local schools to do that is to empower and expect them to engage with their broader communities in defining goals that go well beyond what's mandated and tested by the state. Community inclusion in assessment decisions is one route to potentially very dramatic shifts in what we see as possible for students in our schools.

VIII. Conclusion

This paper laid out a description of the core components of current testing policy in Minnesota, analyzed their suitability for their goals, and offered several recommendations (some big-picture, some very specific) to improve the attention to assessment in this state. This conclusion will describe the new system that would result from implementing all of these changes, evaluate how the new system would compare with the current one in meeting the various goals of testing at different levels of our school system, provide a final description of the proposed system's alignment with the principles laid out in Section II, and offers a closing call to action demanding better assessment policy on behalf of Minnesota students.

PUTTING IT ALL TOGETHER

While each of the recommendations in this paper could be implemented more or less independently of one another, they will also complement and reinforce one another when undertaken together. To review, if Minnesota implemented all of these recommendations:

- MCAs that better assess the breadth and depth of student learning will be administered in fifth grade (representing the elementary grade span), eighth grade (representing the middle school grade span), and high school, continuing to provide annual updates about system-, district-, and school-level trends and gaps between student populations in the tested subjects.
- Students will spend less time taking and preparing for standardized tests.
- Additional attention and professional support will flow to classroom assessment, where the most important assessment and feedback for student learning occurs.
- Professional development will improve, incorporating student data and peer support to meaningfully improve teaching skills.
- An increased climate of professional accountability to one's peers and one's community will add to teachers' existing motivation to improve and give them more tools to pursue that improvement.
- Communities will be better informed about students' progress toward a wide range of academic and non-academic goals that matter to them, and will be welcomed, valued, and supported as partners in the pursuit of those goals.

THE STATUS QUO VS. RECOMMENDATIONS

A responsible discussion of this new system must evaluate how it performs relative to the current system in a number of areas. Specifically, Minnesota's current assessment policies provide information for identifying system-, district-, and school-level trends and concerns, teacher accountability and improvement, and supporting instruction. The following table compares and contrasts the current system with this paper's recommendations in each of these areas.

ASSESSMENT PURPOSE	STATUS QUO	PROPOSED SYSTEM
Identify system-, district-, and school-level trends and concerns	• NAEP provides interstate comparisons. • Annual MCAs in grades 3-8 and high school track gaps and changes in math and reading. Annual grade span MCAs track gaps and changes in science. • Many districts supplement the MCAs with additional tests. • Communities not included in setting and pursuing goals.	• NAEP provides interstate comparisons. • Annual grade span MCAs track gaps and changes in math, reading, and science. • New MCAs assess more deeply than current MCAs. • Educators included in decisions about usefulness of additional tests. • Communities welcomed when setting and pursuing goals.
Teacher accountability and improvement	• Student data used in teacher evaluation. • Future possibility of partially linking pay and/or retention to test-based calculations. • Heavy emphasis on outside, "sit-and-get" professional development.	• Student data used in teacher evaluation. • Increased role of peer-based accountability systems in improving teaching quality and removing consistent under-performers. • More targeted, sustained professional development using local experts.
Support instruction	• Little systemic attention to classroom assessment. • Classroom assessment pressured to mimic shallow standardized test format.	• High systemic attention and support for classroom assessment. • Teachers able and free to use a wide range of robust, appropriate assessments.

ALIGNMENT WITH PRINCIPLES

Each section of this paper has identified how its recommendations align with the principles laid out in Section II. This is a summary review of the proposed system's overall alignment.

STUDENT LEARNING

Students deserve a robust curriculum

The proposed system uses fewer and better state tests to remove the pressure the current system exerts to narrow curriculum. The proposed system also encourages the use of a wide range of assessments in the classroom, and includes the community in setting academic and non-academic goals and selecting assessment tools. All of these changes do a better job of ensuring students' access to the robust curriculum they deserve.

Assessment should reflect learning

The proposed system recognizes the importance of classroom assessment as a key tool for measuring student learning and informing changes to curriculum and instruction. It prioritizes assessment as a teaching skill to ensure that assessments accurately demonstrate learning and can be used by students and educators alike to improve learning.

Assessment should measure growth

While current district-selected tests provide some insight into students' growth, the proposed system places additional emphasis on educators' ability to monitor and report growth toward a much wider range of goals.

Testing should take place on an intentional schedule that does not disrupt classroom learning

The proposed system recognizes the need for tracking gaps and changes in proficiency between student populations at the state and district levels as one part of ensuring our focus on educational equity. The use of grade span testing creates an opportunity to invest in better assessments while still ensuring an annual accounting of gaps and changes at the elementary, middle school, and high school levels.

TEACHER PRACTICE AND ACCOUNTABILITY

Students deserve a learning environment that provides timely and meaningful feedback

The current attention to assessment places the greatest focus on state tests and, to some degree, district-selected standardized tests. However, assessment and feedback have the most impact on student learning when they happen in the classroom. The proposed system places more emphasis on supporting educators as they develop assessment-related skills to ensure students receive the education they deserve.

Standardized assessments are only valuable if they provide meaningful data that educators are trained to use

An educator's ability to accurately analyze data and draw meaningful conclusions for improving their practice requires that they have the training to do so. While the current system is scattershot in its attention to this key concern, the proposed system recognizes data analysis as a key professional skill to be supported by improved professional development and other forms of ongoing education.

Student learning is influenced by variables inside and outside of schools and classrooms

Too often, analysis of state- and district-required test scores lacks a thorough accounting of the impact of other variables on student learning. The proposed system emphasizes greater flexibility and adaptability of classroom assessments and is more proactive about engaging with schools' communities to better shape and pursue goals in a way that recognizes those outside variables without using them as an excuse for inaction.

Quality teaching is measured by numerous variables that include but are not limited to student performance on assessments

The proposed system retains the current system's use of student data as one component of teacher evaluation, but recognizes the limitations of certain proposals for which data to use and for what purpose. The proposed system places significant emphasis on evaluation as a tool for meaningfully improving teaching, and encourages educators to use several types of student data in appropriate ways with their peers, supervisors, and community members.

WHY WE MUST DEMAND BETTER

The recommendations here show that it is possible to have a serious conversation about student learning and assessment while respecting educators' professionalism and commitment to students. The current system has been shaped in too many ways by distrust and outright suspicion of educators. Too often, the current system has narrowed what is taught, encouraged schools and districts to purchase simplistic curricula, and diverted attention from some of the most important factors that affect teaching quality and student learning.

We must demand better. We must demand a system that expects teachers to regularly assess and provide high-quality feedback to their students, and one that supports teachers in developing and maintaining the many skills needed to do so. We must demand a system that values quality of state testing over quantity, while addressing the very real crisis of achievement gaps. We must demand a system that empowers schools and their surrounding communities to work together to define goals that include, but also transcend, those established by the state, and to work together in pursuit of those goals.

The current system demonstrates the inherent flaw of a technocratic approach that assumes that test scores, which too often lack real human context, can be used algorithmically to improve what is, at its core, a system of humans working together. Education is a human act, with educators creating an environment and providing the support for students to learn. The way we think about assessments should reflect that human reality. The educational value of assessments is rarely the final score. It's what a teacher writes to a student, or how a peer connects a teacher's actions to students' performance, or how educators and community members share information and learn from each other as they analyze data together.

That's what assessment can be, and what it is in some places in Minnesota. Our current system of testing, however, does not do enough to prioritize students and educators. We must demand better.

Appendix A: Works Cited

Aldeman, C. and Hyslop, A. (2015). "Grade-span accountability is a bad idea: Just ask CAP and the AFT." Education Next website. <http://educationnext.org/grade-span-accountability-bad-idea-just-ask-cap-aft/>

American Statistical Association. (2014). "ASA statement on using value-added models for educational assessment," American Statistical Association. <https://www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf>

Anderman, E.M., Gimbert, B., O'Connell, A.A., and Riegel, L. (2014). "Approaches to academic growth assessment." British Journal of Educational Psychology.

Au, W. (2011). "Teaching under the New Taylorism: High-stakes testing and the standardization of the 21st Century curriculum." Journal of Curriculum Studies, 43(1), p. 25-45.

Berliner, D. (2011). "Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows." Cambridge Journal of Education, 41(3), p. 287-302.

Berliner, D.C. and Nichols, S.L. (2007). Collateral damage: How high-stakes testing corrupts America's schools. Cambridge, MA: Harvard Education Press.

Birman, B., Bohrnstedt, G., Hannaway, J., O'Day, J., Osher, D., Phillips, G., Salinger, T. (2013). "Three decades of education reform: Are we still 'A Nation at Risk?'" American Institutes of Research. <http://www.air.org/resource/three-decades-education-reform-are-we-still-nation-risk>

Black, P., Harrison, C., Marshall, B., Wiliam, D., & Lee, C. (2003). Assessment for Learning: Putting it into Practice. Buckingham: Open University Press.

Blank, M. (2009). "Community schools and Promise Neighborhoods." Coalition for Community Schools at the Institute for Educational Leadership. <http://www.communityschools.org/assets/1/AssetManager/ PromiseNeighborhoods_CommunitySchools.pdf>

Blank, M.J., Melaville, A., and Shah, B.P. (2003). "Making the difference: Research and practice in community schools." Coalition for Community Schools at the Institute for Educational Leadership.

Butrymowicz, S. and Garland, S. (2012). "How New York City's value-added model compares to what other districts, states are doing," The Hechinger Report. <http://hechingerreport.org/how-new-york-citys-value-added-model-compares-to-what-other-districts-states-are-doing/>

Cassellius, B. February 2, 2015 letter to legislators. Digital copy viewable at <https://s3.amazonaws.com/s3.documentcloud.org/documents/1683988/testing-reduction-advisory-group-cover-letter.pdf>

Chingos, M.M. and West, M.R. (2015). "Why annual statewide testing is critical to judging school quality." Brookings Institution website. <http://www.brookings.edu/research/papers/2015/01/20-chalkboard-annual-testing-chingos-west>

Darling-Hammond, L. (2010). The flat world and education:How America's commitment to equity will determine our future. New York: Teachers College Press.

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., and Yoon, K.S. (2001). "What makes professional development effective? Results from a national sample of teachers." American Educational Research Journal, 38(4), p. 915-945.

Glazerman, S.M. and Potamites, L. (2011). "False performance gains: A critique of successive cohort indicators." Mathematica working paper. <http://www.mathematica-mpr.com/~/media/publications/PDFs/Education/False_Perf.pdf>

Goldhaber, D.D., Brewer, D.J., and Anderson, D.J. (1999). "A three-way error components analysis of educational productivity." Education Economics, 7(3), p. 199-208.

Green, E. (2014). Building a Better Teacher: How Teaching Works (and How to Teach It to Everyone). W.W. Norton & Co: New York, NY.

Harlen, W., and James, M. (1997). "Assessment and learning: Differences and relationships between formative and summative assessment." Assessment in Education: Principles, Policy & Practice, 4(3), p. 365-379.

Hattie, J., and Timperley, H. (2007). "The power of feedback." Review of Educational Research, 77(1), p. 81-112.

Langberg, J., Qureshi, T., and Deas, E. (2013). "Community-based accountability: Best practices for school officials." Poverty & Race Research Action Council, 22(2), p. 9-12.

Marsh, J.A., Springer, M.G., McCaffrey, D.F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., Peng, A. (2011). "A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report," The Rand Corporation. <http://www.rand.org/pubs/monographs/MG1114.html>

Minnesota Department of Education. (2012). "Feedback: Testing in districts," Minnesota Department of Education.

Minnesota Department of Education. (2014). "Reading test specifications for MCA-III, grades 3-8 and 10." Accessible via <http://education.state.mn.us/MDE/EdExc/Testing/TestSpec/>

Minnesota Department of Education. "Testing information." <http://education.state.mn.us/MDE/JustParent/ TestReq/index.html>

Minnesota State Demographic Center. (2013). "In the shadow of the Boomers: Minnesota's labor force outlook." <http://mn.gov/admin/images/in-the-shadow-of-the-boomers-labor-force-outlook-msdc-dec2013.pdf>

Moon, T.R. (2005). "The role of assessment in differentiation." Theory Into Practice, 44(3), p. 226-233.

National Center for Education Statistics. (2014). "NAEP overview," National Center for Education Statistics website. <http://nces.ed.gov/nationsreportcard/about/>

Nelson, H. (2013) "Testing more, teaching less: What America's obsession with student costing costs in money and lost instructional time," American Federation of Teachers.

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 et seq. (West 2003)

Nye, B., Konstantopoulos, S., and Heges, L.V. (2004). "How large are teacher effects?" Educational Evaluation and Policy Analysis, 26(3), p. 237-257.

Papay, J.P. and Johnson, S.M. (2012). "Is PAR a good investment? Understanding the costs and benefits of teacher peer assistance and review programs." Educational Policy, 26(5), p. 696-729.

Post, T. (2013). "New evaluations put Minnesota's teachers to the test." The Hechinger Report. <http://hechingerreport.org/new-evaluations-put-minnesota-teachers-to-the-test/>

Ripley, A. (2013). The Smartest Kids in the World, and How They Got That Way. Simon & Schuster: New York, NY.

Rosenshine, B.V. (1986). "Synthesis of research on explicit teaching." Educational Leadership, 43, p. 60-69.

Sanders, M.G. (2003). "Community involvement in schools: From concept to practice." Education and Urban Society, 35(2), p. 161-180.

Talbert, J.E. (2010). "Professional learning communities at the crossroads: How systems hinder or engender change." Second International Handbook of Educational Change; Springer International Handbooks of Education, 23, p. 555-571.

The Education Trust. (2015). "Five problems with grade-span testing." <http://edtrust.org/wp-content/uploads/2013/10/GradeSpanTesting_Jan2015.pdf>

Tucker, M.S. (2011). "Standing on the shoulders of giants: An American agenda for education reform," National Center on Education and the Economy, p. 8. <http://www.ncee.org/wp-content/uploads/2011/05/Standing-on-the-Shoulders-of-Giants-An-American-Agenda-for-Education-Reform.pdf>

United States. National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform : a report to the Nation and the Secretary of Education, United States Department of Education. Washington, D.C.: The Commission.

U.S. Department of Education. (2012) "ESEA flexibility policy document." Accessible at <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>

Wiggins, G. (1990). "The case for authentic assessment." ERIC Digests. ERIC Clearinghouse on Tests, Measurement, and Evaluation; American Institutes for Research.


epic
EDUCATOR POLICY
INNOVATION CENTER