

## *Background and Rationale for High-Stakes Testing and Accountability Positions*

### Introduction

For nearly three decades, life in public school classrooms in Ohio and across the nation has been transformed by **high-stakes testing**, the use of **standardized tests** to hold students, teachers, and school districts accountable for all students achieving at high levels. In this approach to securing high-quality education, individual and group test performance that fails to meet expectations triggers negative consequences for students, teachers, schools, and school districts. The theory behind test-based accountability is that by demanding all students to meet testing goals, high-stakes tests will motivate students and teachers to work harder, learning will improve, and achievement differences will narrow.

### The Rise of Test-Based Accountability

The Ohio Constitution requires a “thorough and efficient” system of public schools that are free and open to all residents of the state, making education a fundamental responsibility of state government. Education is delivered by local school districts with oversight provided by locally elected boards of education to ensure that the public interest is well served. Taxes raised at the state and local level provide the bulk of funding.

The Elementary and Secondary Education Act, passed in 1965 gave the federal government an important role in advancing equity in public schools. Title 1 of the law provides extra resources to school districts to meet the greater educational costs associated with special needs populations, which are not equally distributed among all school districts. These resources, referred to as **compensatory education**, remain a federal responsibility and were reauthorized in 2002 as No Child Left Behind (NCLB) and again in 2015 as Every Student Succeeds Act (ESSA).

Historically accountability for education quality is a shared responsibility of state government and local boards of education and administrators. Locally elected school boards provide close-up monitoring of how well schools are serving children and meeting community expectations. School administrators evaluate teachers and the quality of instruction and services available to students. Voters help make schools accountable by approving local school funding, electing boards of education, and directly addressing school leaders and teachers with their concerns. A key goal of oversight is to make sure public funds are well spent and that the public schools provide the quality of education promised to all of its citizens. Up until the mid-1980s, states used an “accreditation” approach to accountability that defines inputs to education such as the teaching skills and curriculum content that meet quality standards (Mikulecky & Christie, 2014).

Changes in the U.S. economy caused by a shrinking manufacturing base and the rise of information technology and globalization in the late 20<sup>th</sup> century inspired business and political leaders to advocate for dramatic changes in education (Kirsten, 2002; Ravitch, 2013). The Business Roundtable proposed that all high school graduates have the

higher-level skills needed for the new economy. This led policymakers to inject the business practice of measuring outcomes as the way to improve results. Testing student achievement, a measurable outcome of education, and holding educators and students accountable for those outcomes became the means to improve education.

In 2002, the federal government mandated high-stakes testing as a condition of federal funding under No Child Left Behind, making this a national strategy. The federal law required states to test all students in grades three through eight and once in high school in language arts and math and periodically in science. All students were expected to be “proficient” by 2014. Test data for every school, district, and state had to be reported for subgroups within the school including those defined by race, ethnicity, income, disabilities, English proficiency, and immigration status. Each state was expected to develop standards, tests, and implementation plans to carry out the testing provision. They also defined the consequences of failure to make adequate progress. Child advocates were attracted to the goal of universal success, and the focus on narrowing the achievement gap, which is often blamed on low expectations for some students.

The policy, however, was implemented without being field-tested (Koretz, 2017). Despite the lack of evidence that standardized testing linked to high-stakes consequences would improve student achievement, this approach became the strategy for advancing improvements in school quality. States created their own systems to comply with federal law and soon added more high-stakes requirements. Today, test-based accountability remains the centerpiece of federal and Ohio education policy to improve education.

## High-Stakes Testing Requirements in Ohio

Ohio law mandates that standardized tests aligned with Ohio’s curricular content standards be used to measure student achievement. Test scores play a prominent role in deciding student advancement, evaluating teacher effectiveness, and rating the quality of education in individual buildings, and school districts. In 2018, tests in Ohio are used for these high-stakes decisions:

- Third grade guarantee - Based solely on their performance on the third-grade reading test, individual students can be held back from fourth grade.
- High school graduation - Students who have successfully completed all high school course requirements must also earn enough points on state designed end-of-course tests to receive their diplomas.
- Teacher evaluation - Student test scores account for 50 percent of a teacher’s evaluation and can affect their pay and even put their jobs in jeopardy.
- School rating – Individual schools are ranked on a scale of “A” to “F.” A low rating requires state-mandated improvement plans and after three years triggers the **EdChoice** voucher program, where students residing in that school’s attendance zone are eligible for vouchers to attend parochial or private schools. Voucher funds are subtracted from the state share to the school district, thus diverting resources from low-rated schools and weakening their capacity for improvement.

- District rating – School districts are ranked on a scale of “A” to “F” based almost exclusively on test scores.
- State Takeover – If districts receive an “F” rating for three consecutive years, the local board of education loses its authority and an **academic distress commission (ADC)** appointed by the state assumes control. Youngstown and Lorain are operating under ADCs and East Cleveland was assigned an ADC just this year.

## Limitations of Test-Based Accountability

Ohio’s accountability system carries significant negative consequences for children, teachers, schools, and communities when test results fall short. Those consequences make it important that the system for assigning them is valid and reliable. Ensuring that Ohio’s children are afforded a quality education is desirable, but the test-based accountability approach has limitations that stem from (1) the validity of the theory of change, (2) the misuse of standardized tests, and (3) the impact of high-stakes testing on schools, students, and communities.

### 1. Validity of the theory of change

Test-based accountability defines a test performance level that all students should achieve and then measures progress toward that goal. It assumes that attaching consequences to test performance will motivate teachers to ensure their students learn more. For this to be effective and fair, the expected outcome must be realistic and attainable, and teachers must have the power to deliver those results. One criticism of this approach is that policymakers set high expectations for achievement without evidence that it was attainable. No Child Left Behind set an aspirational goal, not a realistic one (Ravitch, 2013).

Accountability makes teachers responsible for student learning, even though many other factors that teachers cannot control affect student achievement. It ignores that individual differences affect learning, that education opportunities are not equal in every school, and that out-of-school factors such as housing stability, trauma, health and nutrition, parental engagement, and enrichment experiences also affect educational attainment.

Many education experts questioned how measuring achievement could produce results if nothing was done to address the capacity of schools to raise achievement or the life circumstances and out-of-school factors that affect student learning. Others questioned that the strategy was expecting schools to overcome the effects of poverty. As George Wood (2004, p.49), principal of Federal Hocking High School in Stewart, Ohio, observed, “To hold schools accountable without appropriate support is worse than a farce, it is a tragedy.”

By ignoring the well-documented influence of resources and out of school factors on student achievement, accountability distributes sanctions in ways that furthers inequality.

### 2. The misuse of standardized tests

Standardized tests are the core ingredient of test-based accountability, but testing experts have repeatedly expressed concern that basing high-stakes decisions on test results is a misuse of standardized tests. For example, at the time NCLB was considered and enacted,

the Center on Education Policy issued a report that explained the value of standardized tests as a resource for making comparisons about student performance across locations or between and within groups, for diagnosing skills for individuals or groups, and for informing instruction, but warned that they are too imprecise for making serious decisions (Kober, 2002, Oct). Education historian Diane Ravitch (2010, p. 150) summed up the problem, “Policy decisions that were momentous for students and educators came down from elected officials who did not understand the limitations of testing.”

Here are key reasons why standardized tests are not appropriate for accountability:

- Standardized tests are imprecise measures of student achievement. Standardized test scores can change from one day to the next. They are an estimate of what a student has learned. They are informative but not sufficient for making important judgments or decisions (Kober, 2002; Koretz, 2017).
- Standardized tests measure what students know but not why they know it. It is inappropriate to use standardized tests to judge teacher effectiveness or school quality. Daniel Koretz (2017, p. 148) warns, “tests are very useful for describing what students know. On their own, tests simply aren’t sufficient to capture why they know it.”
- High-stakes tests narrow the purpose of education. Test-based accountability focuses on the most easily measured content. By stressing what is measurable, it ignores other equally important purposes of education that are not easily measured: shaping character, developing sound minds and healthy bodies, and forming citizens for a democracy (Ravitch, 2010).
- High-stake tests undermine learning by emphasizing test scores. When test scores are used to measure school quality, it provides a powerful incentive for educators to find short cuts to raise test scores rather than improve instruction. As Diane Ravitch (2010, p. 154) explains, “When the purpose of testing is informational and diagnostic, there is no reason for teachers and administrators to alter the results except through improved instruction. But when the purpose of testing is accountability, then teachers and administrators understand that there are real consequences if the scores in their classroom or their school change.”

### **3. Impact on the classroom, students, and communities**

A danger of high-stakes testing is that test scores become the goal of the teaching process. When this happens, “[...]tests] lose their value as indicators of educational status and distort the education process in undesirable ways” (Koretz, 2017, p. 38). Giving tests high-stake consequences inspires educators to game the system to ensure the highest possible test performance. As a result, better test scores are not evidence of more learning or improved education, but the need for high scores can undermine education quality.

Here are some ways that educators respond to the use of test results to make decisions that are harmful to education quality:

- More class time is spent on content for subjects likely to be tested. This narrows rather than enriches the curriculum.

- Teachers focus on specific groups of students for whom the expected scores are most attainable, undermining success for all students.
- Alternatively, students who are the most vulnerable are likely to spend time on test preparation at the expense of exploring a rich curriculum.
- Time spent on test-taking skills and practice tests detract from instructional time and communicate that learning is test success. Test-taking strategies may increase the chances of a correct answer but will not improve learning.
- Schools have given up field trips, recess, play-based learning, and other types of engaged instruction to make more time for test preparation. The school calendar is designed to maximize class time before testing.
- Schools devote precious resources to administer tests and raise scores. These resources include valuable instructional time, funds for testing materials and testing technology, and staff time for logistical test preparation and proctor training. Many school districts have found it necessary to hire a testing coordinator for each school. Students not participating in certain tests are often left to do busy work under substitute teachers.

Also, because teacher evaluations are based on test scores, more tests are created to evaluate teachers not teaching in areas mandated by federal law, taking away more time from instruction and excitement away from learning.

The correlation between income and test performance has been consistent over time but this reality is not factored into test-based accountability. As a result, the focus on test prep over enriched learning occurs predominantly in schools serving the most vulnerable students because they have the farthest to go to meet testing expectations. Similarly, negative consequences are most frequently assigned to these students and the schools that serve them. When test consequences fail to account for out-of-school factors or inadequate school resources, the system disadvantages some students and favors others. It can mask good teaching in difficult situations, and bad teaching in places where there are fewer challenges. This too works against equalizing outcomes by discouraging great teachers from serving students whose test performance will define them as failures (Darling-Hammond, 2004).

Furthermore, whole districts and even communities have been impacted by accountability based on test results. Real estate values and the desirability of a community depend on confidence in the public schools. A negative district rating can hurt the vitality of a community and prospects of voters approving additional education funds. The authorization of state takeovers in 2015 by HB 70 has put three school districts under ADCs: Youngstown, in 2015; Lorain, in 2017; and East Cleveland, in 2018. These are all high-poverty school districts. In addition to taking control of the schools out of the hands of the local population, HB70 accelerates programs for “school choice,” further compromising school district resources. Test performance has not improved despite state takeover.

## Failure of High-Stakes Accountability to Improve Education

After years of experience with high-stakes testing, outcomes have not changed in any marked and sustained way (Koretz, 2017; Barnum, 2018; Greene, 2018). Achievement gaps have not narrowed. Koretz calls test-driven school improvement “a failure.”

In Ohio, test scores remain flat. The headline for a *Plain Dealer* article reporting on 2013 test results captures it well, “State report card shows: student performance is related to income and poverty levels – yet again” (O’Donnell, 2013). In the 2018 test results, of the 28 districts receiving “A” grades, 23 were in Ohio’s wealthiest suburbs and five were in small towns. Fourteen districts received an “F” including the state’s largest urban districts and three poor, segregated first ring suburbs (Resseger, 2018, Sept. 14).

High-stakes testing has taken the life out of classrooms, extinguishing teacher enthusiasm for teaching and students’ love of learning. In response to the time spent on testing and the pressure to perform, students frequently exhibit anxiety, apathy, and disengagement. Many classroom teachers have observed that this causes struggling students to give up, discourages attendance, and encourages some students to drop out.

It is harmful to label students, teachers, schools, and communities as successes or failures based on test results. These judgments can destroy public confidence and investment in public education, weaken local community institutions, and put the reputation and viability of diverse communities at risk. High-stakes testing has fallen short of its promise to improve public education. It is undercutting this goal while students, schools, teachers, and communities suffer.

## Proposed Positions on High-Stakes Testing and Accountability

Ohio’s citizens and policymakers want students in our state to receive an effective education that inspires love of learning, supports the unique needs and interests of each student, and prepares them for life. Test-based accountability too frequently works against these goals. It is neither fair, nor are judgments tied to test results an appropriate use of standardized tests. The negative effects on life in our schools is detrimental to the long-term success of public education.

Members of the League of Women Voters of Pennsylvania (LWVPA) and of the League of Women Voters of Texas (LWVTX) (see “League of Women Voters Study Resources” below) have questioned high-stakes testing and adopted positions to address the appropriate use of standardized tests and their role in school accountability. We share their concern. The negative effects on children, authentic learning, and teacher efficacy and the unfair labeling of schools as failures or successes because of whom they serve motivated our group to review the validity and fairness of how tests are used and the judgments made based on results. We believe that the positions of LWVPA, especially, capture our concerns and would permit the LWV of Ohio to advocate for education policies that can regain balance in the way we pursue quality in Ohio’s public schools.

## LEAGUE OF WOMEN VOTERS STUDY RESOURCES

League of Women Voters (LWV) of Pennsylvania positions “High-Stakes Testing and Assessment” (2015)

<http://www.palwv.org/wp-content/uploads/2014/04/Education-Position-8-2015.pdf>

LWV Pennsylvania study (2015): “Education Study Guide High-stake Testing”

<http://lwvpgh.org/files/HST-study-guide.pdf>

League of Women Voters of Texas positions “Public School Testing and Accountability” (2008)

<https://my.lwv.org/texas/position/public-school-testing-and-accountability>.

LWV Texas study (2007): “Mandated Achievement Testing in the Public Schools of Texas Facts & Issues.” (Link from above URL)

## BIBLIOGRAPHY

Barnum, Matt. (2018, Aug. 13). “Pressuring schools to raise test scores got diminishing returns, new study of No Child Left Behind finds.” *Chalktalk*.  
<https://chalkbeat.org/posts/us/2018/08/13/pressuring-schools-to-raise-test-scores-got-diminishing-returns-new-study-of-no-child-left-behind-finds/>

Darling-Hammond, Linda. (2014). “From ‘Separate but Equal’ to ‘No Child Left Behind:’ The Collision of New Standards and Old Inequalities.” In D. Meier and G. Wood (eds.). *Many Children Left Behind* (pp. 3-32). Boston, MA: Beacon Press.

Greene, Peter. (2018, Sept. 20). “Is the big standardized test a big standardized flop.” *Forbes*.  
<https://www.forbes.com/sites/petergreene/2018/09/20/is-the-big-standardized-test-a-big-standardized-flop/#4702cf974937>

Hess, Frederick. (2018, Sept. 18). “How much do rising test scores tell us about a school?” *Forbes*.  
<https://www.forbes.com/sites/frederickhess/2018/09/18/how-much-do-rising-test-scores-tell-us-about-a-school/#36e4812d22e8>

Jennings, Jack; Diane Stark Rentner & Nancy Kober. (2002, Jan.). “A New Federal Role in Education.” Washington, DC: Center on Education Policy.  
<https://www.cep-dc.org/data/searchquery.cfm?search=Yes>

Kober, Nancy. (2002, Oct.). “What tests can and cannot tell us.” *TestTalk for Leaders*. Washington, DC: Center on Education Policy.  
<https://www.cep-dc.org/displayDocument.cfm?DocumentID=257>

Koretz, Daniel. (2017). *The Testing Charade*. Chicago, IL: University of Chicago Press.

Larson, Kirstin, ERIC Clearinghouse on Educational Management (2002, Summer). “Corporate Involvement in School Reform.” *Policy Report*, 3. Eugene, OR: ERIC Clearinghouse on Educational Management, College of Education, University of Oregon.  
<http://hdl.handle.net/1794/3463>

Mikulecky, Marga & Kathy Christie. (2014). *What Parents and Experts Say States Should Consider to Make Accountability Systems Meaningful*. Denver: Education Commission of the States.

<http://www.ecs.org/docs/rating-states,grading-schools.pdf>

O'Donnell, Patrick. (2013, Sept. 23). "State report cards show student performance is related to income and poverty levels – yet again," *Cleveland.com*.

[https://www.cleveland.com/metro/index.ssf/2013/09/state\\_report\\_cards\\_showstudent.html](https://www.cleveland.com/metro/index.ssf/2013/09/state_report_cards_showstudent.html)

Ravitch, Diane. (2010). *The Death and Life of the Great American School*. New York, NY: Basic Books.

Ravitch, Diane. (2013). *Reign of Error*, New York, NY: Knopf.

Resseger, Jan. (2015, Dec.7). "Ohio Research Proves Yet Again that Test Scores Measure Primarily Family Income." *janresseger*.

<https://janresseger.wordpress.com/2015/12/07/ohio-researcher-proves-yet-again-that-test-scores-measure-primarily-family-income/>

Resseger, Jan. (2018, Sept. 14). "Ohio Releases 2018 School Report Cards, Brands Poorest School Districts with Fs." *janresseger*.

<https://janresseger.wordpress.com/2018/11/01/17869/>

Resseger, Jan. (2018, Oct. 3). "Schools Serving Very Poor Children Need Financial Assistance. Instead Ohio Beats Up on Them." *janresseger*.

<https://janresseger.wordpress.com/2018/11/01/17869/>

Wood, George. (2004). "A View from the Field; NCLB's Effects on Classrooms and Schools." In D. Meier and G. Wood (Eds.). *Many Children Left Behind* (pp. 33-52). Boston MA: Beacon Press.

## GLOSSARY

Academic distress commission (ADC) – Ohio law requires the Superintendent of Public Instruction to establish an Academic Distress Commission for each school district that has been declared to be in academic emergency. (Ohio Dept. of Education)

Compensatory education - supplementary programs or services designed to help children at risk of cognitive impairment and low educational achievement succeed. (Wikipedia)

EdChoice - The Educational Choice Scholarship Program provides students from underperforming public schools the opportunity to attend participating private schools. The program also provides low income students who are entering kindergarten through fifth grade the same opportunity. The program provides up to 60,000 EdChoice scholarships to eligible students. (Ohio Department of Education)

High-stakes test - any test used to make important decisions about students, educators, schools, or districts, most commonly for the purpose of accountability—i.e., the attempt by federal, state, or local government agencies and school administrators to ensure that students are enrolled in effective schools and being taught by effective teachers. In general, "high stakes" means that test scores are used to determine punishments (such as sanctions,

penalties, funding reductions, negative publicity), accolades (awards, public celebration, positive publicity), advancement (grade promotion or graduation for students), or compensation (salary increases or bonuses for administrators and teachers). (edglossary.org)

Standardized test - test that (1) requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same way, and that (2) is scored in a “standard” or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. While different types of tests and assessments may be “standardized” in this way, the term is primarily associated with large-scale tests administered to large populations of students.... (edglossary.org)

Theory of change - a comprehensive description and illustration of how and why a desired change is expected to happen in a particular context. It is focused in particular on mapping out or “filling in” what has been described as the “missing middle” between what a program or change initiative does (its activities or interventions) and how these lead to desired goals being achieved. It does this by first identifying the desired long-term goals and then works back from these to identify all the conditions (outcomes) that must be in place (and how these related to one another causally) for the goals to occur. These are all mapped out in an Outcomes Framework. (theoryofchange.org)