

The United States: School Choice and Test-Based Accountability

Kevin G. Welner

Abstract This is an unusual time in the USA for policy concerning primary and secondary education—kindergarten through 12th grade. The forces shaping K-12 policy are remarkably aligned with both major political parties devoted to two fundamental approaches: test-based accountability and school choice. While these lawmakers differ over details, including the proper role of the federal government, there is little disagreement regarding reliance on these basic approaches. While individual states and school districts have embarked on enough different reforms so as to decorate this remarkably aligned political landscape with a variety of interesting gardens worthy of notice, this chapter focuses on explaining the history and current import of the two dominant policies.

Keywords Test-based accountability • School choice • US policy

This is an unusual time in the USA for policy concerning primary and secondary education—kindergarten through 12th grade. The forces shaping k-12 policy are remarkably aligned with both major political parties devoted to two fundamental approaches: test-based accountability and school choice. While these lawmakers differ over details, there is little disagreement regarding reliance on these basic approaches. Yet even within such an aligned political landscape, individual states and school districts have embarked on enough different reforms so as to decorate the landscape with a variety of interesting gardens worthy of notice.

This chapter therefore describes the nature, appeal, and consequences of the two primary strands of policy even while making note of the complexities and diversity co-existing—and often standing in tension—with the dominant policy agenda. The first section sets forth the basic legal structure for public schooling in the USA. This is followed by a very brief presentation of historical eras of education policy and

K.G. Welner (✉)
University of Colorado Boulder, School of Education, USA
e-mail: kevin.welner@Colorado.edu

then an overview of recent changes to that policy and to underlying principles. The bulk of the chapter then focuses on test-based accountability and school choice, the policies that now dominate schooling in the USA.

1 The Legal Structure of, and Recent Changes to, US Educational Policy-Making

According to the National Center for Educational Statistics (NCES), schools in the USA enroll 55 million students in grades prekindergarten through 12 (aged approximately four through 18), with 50 million of those students enrolled in public schools (National Center for Educational Statistics n.d.a). Most students persist through graduation; students categorized as Asian graduated at the highest rate (79 %), while Black and Latino students had graduation rates of 60 % and 58 %, respectively (Chen 2012). Nationally, almost half (48 %) of all public school students are from low-income families. Over one-eighth (13 %) are identified as having disabilities, and a growing number—again 13 %—are Limited English Proficient (National Center for Educational Statistics n.d.b).

Across the USA, there are almost 93,000 public schools serving grades kindergarten through 12, and these schools “will spend about \$571 billion for the 2012–2013 school year. On average, the current expenditure per student is projected at \$11,467 for this school year” (National Center for Educational Statistics n.d.c). Because of great variation in the cost of living, in needs such as transportation, in student poverty levels, and in a given state’s ability and level of commitment to funding education, state-level annual per pupil spending averages vary from \$6,064 (in Utah) to \$18,618 (in New York) (National Center for Educational Statistics n.d.a).¹

Meaningful international spending comparisons are difficult, since the US has an employment-based system of health care and retirement pensions. For instance, the \$18,618 figure for New York includes \$4,658 just for employee benefits—which is in addition to \$10,895 for salaries and wages. Nonetheless, comparisons are available, including an NCES determination that “In 2008, the United States spent \$10,995 per student on elementary and secondary education, which was 35 % higher than the OECD average of \$8,169” (National Center for Educational Statistics n.d.d).

US spending figures are based on the addition of spending at federal, state, and local levels, reflecting a comparable distribution of authority and responsibility. Education of k–12 students in the USA is primarily a state-level responsibility, but the federal government has made inroads over the past half century—and particularly over the past decade. The following is a brief discussion of the federal and state roles in policy-making as well as the authority exercised at other levels.

As a legal structure, the system has many layers of rule-making. At the ground level are individual classrooms, where teachers were formerly able to exercise a great

¹ See Table 8 at http://www2.census.gov/govs/school/elsec10_sttables.xls, for school year 2009–2010 (most recent data).

deal of discretion. This was true in the 1970s and 1980s but, as discussed below, has changed over the past couple decades. Beyond the classroom level, different school systems distribute authority in a variety of ways. In high schools, for instance, additional decision-making sometimes takes place at the department level, where there may be a mathematics or sciences or English chair who is vested with some authority over instruction and supervision. At the school level, the principal generally exercises considerable personnel and budgetary authority as well as some authority over curriculum and instruction. District-level governance differs from state to state, but there is typically a school board—which is usually elected by community members—which hires a district superintendent, who in turn hires other district staff and school leaders. The school board and superintendent are usually vested with ultimate authority over district-level policies, including key budgetary decisions and broad policy-making.

Recognizing these more local levels of decision-making is important because the educational system is loosely coupled (Weick 1976). Dictates from higher levels of the political system do not make their way down to classrooms with perfect fidelity; for better or worse, they are often watered down and adapted to localized needs and preferences (Welner 2001).

But an increasing number of critical educational policies—arguably most major policies—are now determined at the state and federal levels. At each of these two levels, there are statutes, constitutional provisions, and departments of education with rule-making authority. Because the USA was created as a federation of states, the federal government is akin to national governments in most countries. Yet the US Constitution, which is the ultimate law of the land, includes nothing specific about education. In fact, throughout most of the nation's history, the federal government played only a minor role in k–12 education. This limited role was focused on efforts such as data gathering and research and then, starting in the 1960s, civil rights and equity concerns. Courts, at both the federal level and in the states, have also played a key role over the years, as they were called upon to interpret and give teeth to protections set forth in constitutional and statutory provisions (Welner et al. 2010).

Until very recently, the most important decisions were largely left to the discretion of the legislators and governors in each of the 50 states, who created a patchwork of different systems. These various state systems certainly had many commonalities, which has been important because families often move from state to state. Such similarities facilitated such things as college applications across state lines as well as a relative continuity of curriculum progression for, e.g., a 10-year-old child who moves from New York to Pennsylvania.

But many elements were also different, including the ages of mandatory schooling, the required qualifications of teachers, and the treatment of students with special needs or with a first language other than English. To a considerable extent, these variations still remain, but federal interventions have resulted in greater consistency. Most recently, and as discussed below, the US Department of Education has pressured states to adopt academic standards to ensure that all students are “college and career ready”—in furtherance of the so-called Common Core State Standards movement. Note that although these are designated “academic standards” rather than “curriculum standards”, almost all states are also adopting one of two high-stakes assessments (also discussed below), which are very likely to drive more uniformity in curriculum.

This dynamic has, over the past couple decades, yielded an uneasy and ever-changing relationship between the states and the federal government. During that period of time, the nation has seen the two main political parties come into alignment on education issues. That is, this alignment is a relatively recent phenomenon. From the 1960s through the 1980s, the parties staked out substantially different positions. Democrats pushed for an equity-focused and expanded federal role. The main civil rights laws came out of this effort, as did the Elementary and Secondary Education Act (ESEA), the main source of federal educational funding.² Most of ESEA's funding is allocated through Title I of the Act and is designed to provide compensatory resources for students in low-income communities. Republicans, in turn, pushed back against many of these efforts. They contended that the federal government should not interfere with the states' historically dominant role in education.

This all changed with a series of bipartisan federal laws, including the two most recent reauthorizations of ESEA: the Improving America's Schools Act of 1994 and the No Child Left Behind Act (NCLB) of 2001. NCLB in particular illustrates the new landscape for US educational policy. Title I compensatory funding was increased, which pleased most Democrats. And, as discussed more fully below, Republican legislators as well as President George W. Bush were exceptionally keen to install test-based accountability systems that made extraordinarily ambitious demands on the public schools.

2 Shifting Principles

This abbreviated history hints at changes over time in the nation's dominant policy approaches to schooling. When the twentieth century began, education was still largely for the elite. Many states did not even require children to attend elementary schools. Compulsory education laws, as well as laws forbidding child labor, were part of the so-called progressive-era reforms during the first 30–40 years of the century. The universal schooling that arose out of these reforms was, however, highly unequal. Wealthy, white, male students were commonly prepared for college, but the same opportunities were regularly denied to Native American, Latino, Asian American or African American students, as well as to female students, poor students, those with special educational needs, or immigrant students who did not speak English (Tyack 1974). Slow progress was made in some of these areas, but it was not until 10 years after the US Supreme Court's 1954 decision in *Brown v. Board of Education*, which prohibited intentional racial segregation in k–12 schools, that change began to occur at a more rapid pace.

The era of equity reform in education, which lasted from approximately 1964 to approximately 1980, saw enormous strides for all these groups of students. This era

²Notwithstanding the differences between the parties, many of these efforts received substantial bipartisan support.

correspondingly saw enhanced outcomes, with less segregation, greater access, and improved test scores and graduation rates (see Darling-Hammond 2007).

To understand what has happened since, however, it is important first to understand the underlying principles behind the nation's different reform approaches. "Almost every educational policy that finds support in the United States advances one or more of the following ideals: excellence, accountability, equity, innovation, social cohesion, security, efficiency, choice, and meritocracy" (Welner and Oakes 2007, p. 93). Sometimes these goals reinforce one another, but sometimes they exist in tension.

One tension in particular is important to tease out. In the years from approximately 1964–1980, equity and fairness were pursued through policies broadly increasing access to opportunities to learn. Examples include funding and resource equity, civil rights and anti-discrimination protections, reforms of special education policy and language access policy, and class-size reduction. In the years since, however, the policies put forward to enhance equity and fairness have taken a different approach. They have focused on increasing access to individual choice, along with test-based accountability systems that publicly label schools, teachers, and others as succeeding or failing. That is, the nation saw a policy shift from one addressing structural inequalities at a societal level plus large-scale compensatory programs that emphasize inputs to one focused on enhancing individual choice options within a more market-based and deregulated system that emphasizes outcome measures.³

This shift has not been complete, and the past decades have also seen additional, and sometimes countervailing, reform agendas. These include ongoing funding equity lawsuits, policies designed to increase safety and access for gay and lesbian students, class-size reduction and teacher quality initiatives, and early reading intervention programs, as well as pushes for early childhood education and for the reduction or elimination of curriculum tracking. But the movement toward policies emphasizing testing and choice is stark, clear, and overwhelming.

Advocates for this shift spotlight low-income families living in communities with poor schools that serve as little more than "dropout factories"—with few graduates and even fewer students who matriculate to college. The status quo for educational results in such communities is unacceptable. Accordingly, severe accountability rules have a gut-level appeal, particularly at a time when state budgets are strapped. That is, if we cannot invest in schooling, at least we can demand more of the students, teachers, and principals in those schools. There is also a gut-level appeal to school choice policies that give some of these families the opportunity to seek out and enroll in schools that sometimes have more resources and better outcomes. Examples of such policies include charter schools, open enrollment, vouchers to attend private school, tax credits for private schooling, protection of families that choose to home-schooling, and (most recently) online or cyber schooling.

³Note that while Title I funding has continued to increase and while this money continues to be spent overwhelmingly on compensatory needs in low-income communities, the policy conditions attached to this spending now emphasize test-based accountability and, to a lesser extent, school choice.

As discussed below, neither testing policies nor choice policies have found substantial support from research, but they both remain popular among lawmakers. School choice policies in particular have seen unprecedentedly rapid growth over the past decade.

Underlying testing and choice policies are deeper changes to fundamental values and beliefs. In particular, this shift reflects a shift in thinking about the importance or unimportance of larger societal inequalities. The basic concepts of fairness and opportunity have long been deeply embedded in the nation's core values about public education (Oakes et al. 2000). Further, Americans have long looked to public schools as the main institution of social mobility, an idea that can be traced back at least as far as Horace Mann's celebrated call in the mid-nineteenth century for schools to be the "great equalizer" and the "balance wheel of society" (Mann 1848/1868, p. 669).

When a lawmaker is paying attention to larger inequalities, it makes little sense to pursue equity through parental choice. In such a system, parents with less education and less wealth are generally in a relatively worse position to secure advantages for their children (Yettick et al. 2008; see also Wells and Serna 1996). Better educated parents with more resources have more social capital and capacity to "work the system". Accordingly, choice mechanisms will often lead to a situation where the rich get richer and the poor get poorer. While there certainly are disadvantaged families who use choice policies to find better schools for their children, the broad trends are not encouraging (Miron et al. 2012).

3 Choice and Accountability: The Current Trends in US Education Policy

At first glance, school choice policies and test-based accountability policies appear to be separate and distinct. Each one certainly could be pursued in the absence of the other. But as designed and implemented in the USA, the two policies have become intertwined. When schools' test scores are too low, what should be done? Lawmakers are increasingly turning to school choice. How do parents decide whether to choose a school outside the immediate neighborhood? Many look to test scores and the "school grades" based on those test scores. In ways such as these, the two policies have become mutually reinforcing.

3.1 Standards and Test-Based Accountability

Back in the era of equity reform in education, school improvement was primarily driven by inputs. Additional resources, supports, and access were expected to yield gains in outputs, but it was those inputs that garnered the attention of lawmakers. In contrast, the current reform agenda is focused on outputs. In particular, tests are expected to drive school improvement. These policies are often called "accountability

reforms” and are linked to curriculum standards, performance standards, and standardized assessments. Further, as discussed below, these standards-based accountability policies are now being connected to teacher quality policies, grade retention policies⁴, a movement toward a nationwide “common core” set of academic standards, and of course to school choice.

The allure of standardized tests is nothing new in the USA (Gould 1981). They have been used for everything from admission into graduate and professional schools to undergraduate college admissions, to civil service and military appointments, to placement in gifted programs and into different levels (or “tracks”) within schools, and to employment decisions made by private companies. The current testing and accountability push is remarkable in part because of its scope; the amount of testing of k–12 students has increased markedly. But even more remarkable is a key change in use. Prior to 2002 NCLB law, such tests were used to understand the learning needs of, and to make decisions about, the test-takers themselves. They are now being used to make decisions about teachers, principals, schools, and school districts. That is, students’ test scores are being imputed to those around them, to hold those adults and institutions accountable for student learning.

These accountability policies have impacted almost everything we now see in US schools. Decisions about governance, funding, attendance, personnel, curriculum, and instruction have all been transformed. School years even start earlier, so as to squeeze more learning days in, prior to the spring testing period. Keeping in mind the variation between each state’s laws, the systems generally include the following nine elements:

1. Standards that set forth the curriculum at each grade level
2. Standards that set forth what each student should be able to know and do at each grade level
3. Standards that set forth what each teacher should know and be prepared to do
4. Assessments for each grade level
5. Standards (or “cut points”) setting forth proficiency on each assessment or setting forth a system for measuring growth
6. A set of rules for holding teachers accountable for the test scores of students in their classes
7. A set of rules for holding principals accountable for the test scores of students in their schools
8. A set of rules for holding schools accountable for their students’ test scores
9. A set of rules for holding school districts accountable for their students’ test scores

⁴Grade retention and exit (diploma) exams are the only major student-level policies arising out of the recent movement for greater reliance on test scores. While not discussed further below, it should be noted that the retention push has been strong in recent years. As a way to address the importance of early reading skills, Florida and other states have adopted test-based grade retention policies. Student with low scores on the third-grade reading assessment must repeat the third grade. In Florida, these retained students are also provided with intensive reading interventions, which complicates attempts to measure the effectiveness of the grade retention itself. But a great deal of other research concludes that retention is not an effective intervention and, in fact, puts students at a substantially greater risk of later dropping out of school (see studies cited in Moreno 2012).

An additional element calls for these various components to be aligned with one another and with other factors, such as textbooks and teacher preparation programs at college and universities (Smith and O'Day 1990).

Under NCLB, each state was given the discretion to set its own standards, assessments, and cut scores for proficiency, with very little input from the federal government. Each school was, however, required to meet Adequate Yearly Progress (AYP) targets, meaning that they would have to post ever-increasing test scores—up to the point that all students would be proficient by the year 2014 (with several limited exceptions). In fact, the test scores were required to be reported for various disaggregated subgroups (e.g., special needs students, English-language learners, and African American students), and each of these subgroups was also required to show the necessary progress (AYP) toward 100 % proficiency.⁵

When schools failed to clear the NCLB hurdles, they were forced to march through a series of escalating interventions. According to the law, in their first year of “needing improvement”, these schools must provide their students with options for attending nearby, better-performing public schools. The next year, the schools must also allow for a diversion of their Title I funding to pay external service providers for students’ tutoring. The third year, they must implement certain listed “corrective actions”. The fourth year, they must make plans for a change in governance. Then, in the fifth year of “needing improvement”, the school must implement a mandated school restructuring (see discussion in Mathis 2009).

This all gets very complicated; suffice it to say that the Obama administration was aware that more and more schools were swept into this system of escalating sanctions and saw the need to grant waivers to release states from the sanction requirements. Not surprisingly, most states have indeed submitted plans that have been approved by the US Department of Education in exchange for these waivers. Importantly, and controversially, the department did not simply issue waivers; it insisted that these state plans include provisions to promote the administration’s own favored policies—in particular the adoption of “college and career ready standards” plus teacher and principal evaluations linked to students’ academic growth. (Both of these policies are discussed below.) That is, while NCLB’s system technically remained the law of the land, the administration was able to use the waiver process to neuter the AYP sanction system and replace it with a set of different policies—but the replacement policies are again linked to standards and testing.⁶

NCLB requires⁷ that public school students be tested annually in reading and mathematics in grades 3–8, tested twice in the elementary grades in science, and tested in reading, math, and science at least once in grades 10–12. This is much

⁵This requirement of disaggregated subgroup reporting and accountability was one of the elements of NCLB that brought together Democrats and Republicans.

⁶As of July of 2013, eleven states were still operating under the old system, having either not applied for or not been granted a waiver. These include the large states of Texas and California.

⁷The focus herein on NCLB’s accountability provisions. It also includes many other elements, including a provision requiring that teachers be “highly qualified” and teach classes within their area of training.

more testing than had existed in most states before the law, yet the trend in more recent years is still upwards. The Obama administration's policies for the evaluation of teachers and principals call for students' academic growth to be measured in all grades and in all subjects—that is, for all teachers. While the administration cannot directly mandate such policies, it has been successful in coercing states into making the desired policy changes. The conditions imposed on NCLB waivers, noted above, are part of this. Additional leverage has been exercised through competitive grants such as the “Race to the Top” program funded by the so-called stimulus money allocated as part of the American Recovery and Reinvestment Act of 2009.

Academic growth can certainly be measured in ways that do not depend on standardized assessments. But in practice the states have turned to these tests as a means of complying with the administration's demands. They have used two approaches for measuring growth: either value-added modeling (VAM), which attempts to isolate classroom learning effects, or the “Colorado Growth Model”, which is based on student growth percentiles, descriptive measures of relative growth (similar to the children's growth percentile tables used by doctors). Each of these two approaches has problems in terms of measurement and validity, although most published critiques have focused on VAM, which is the approach most states are using (see Baker et al. 2010; Braun 2005; Corcoran 2010; Rothstein 2009).

Each approach also raises serious concerns about the effects of test-driven reform on curriculum and instruction. Perhaps the most important lesson offered by our experience with No Child Left Behind is that schools placed under an incentive system linked to test scores responded by narrowing the curriculum and teaching to the test (Nichols and Berliner 2007). There is no question that high-stakes testing creates powerful incentives and thereby changes behavior, but the key question for lawmakers considering a test-driven reform is whether the exact nature of the incentives will result in beneficial change (Welner 2013).

Implementation is just the beginning of policies that use student test scores as part of teacher and principal evaluation systems. The school districts in Washington DC the state of Tennessee, and a few other jurisdictions have been implementing such systems for the last 2–3 years, but many other states began these systems in 2012–2013 or will begin in 2013–2014. We can expect that empirical analyses of their effects will soon be published.

Test scores have also been used to “grade” schools. Some states expressly attach grades of A–F to schools. In Louisiana, for example, these A–F grades then connect to the state's school choice policies; students attending schools with grades of D or F are given priority to participate in the state's voucher program funding private school tuition. Other notable policies similarly connect assessment results to school choice. As mentioned above, NCLB's escalating sanctions include open enrollment (intra-district public school choice). In addition, one of the listed school restructuring options available to schools that hit the fifth year of “needing improvement” is conversion to a charter school. Similarly, charter conversion is one of four options available under the federal “school improvement grant” program for schools with low test scores and other evidence of low achievement. The new “parent trigger” laws in California and other states also promote charter conversion as an option for

certain low-scoring schools. *What all these laws have in common is a logic model connecting students' test scores to the need for drastic interventions and then turning to school choice as a preferred intervention.*

As noted above, the Obama administration has pressured states to adopt what they call “college and career ready standards”. Owing to the historically limited educational role of the federal government, the administration has kept the development of these standards at arm’s length. Instead, an effort to create “Common Core State Standards” (CCSS) has been led by the National Governors’ Association and the Council of Chief State School Officers, accompanied by fiscal support from the Gates Foundation and other organizations (see Mathis 2012).

The CCSS are set forth at a fairly general level; for example, the “Phonics and Word Recognition” standard for third-grade students (approximately 8 years old) reads in its entirety as follows: “Know and apply grade-level phonics and word analysis skills in decoding words. (a) Identify and know the meaning of the most common prefixes and derivational suffixes; (b) Decode words with common Latin suffixes; (c) Decode multisyllable words; and (d) Read grade-appropriate irregularly spelled words”. The task of putting curricular flesh on these bones is left to the states and, as a practical matter, to the developers of the high-stakes tests based on the CCSS.

A primary impetus for this effort is a concern that many states have been lax in creating rigorous standards and in setting rigorous thresholds for proficiency. Holding students across the nation to the same high standards could, the argument goes, result in the ability to compare results across states—information that is currently supplied by the National Assessment of Educational Progress (NAEP) but that is not provided very well by the hodge podge of state-level assessments. Two national assessment consortia (the Smarter Balanced Assessment Consortium and the Partnership for Assessment of Readiness for College and Careers) are developing computer-based testing⁸ and expect to begin administration of those tests in the 2014–2015 school year. In fact, another anticipated advantage of widespread adoption of the CCSS is the creation of economies of scale for, e.g., private and public organizations that will supply professional development, instructional materials, and standardized testing.

In some ways, this brave new world presents exciting possibilities. But all of the above-described efforts over the past couple decades have been built on the same set of reform assumptions, few if any of which have proven to be correct. With regard to test-based accountability, six main assumptions are as follows:

1. We can design and implement a series of paper-and-pencil (or computer-based) tests that do a good job in capturing what we as a nation think is important for students to know and to do.
2. The results of those assessments can precisely, accurately, and validly distinguish between different students so well that we can make high-stakes decisions about those students.

⁸Smarter Balanced is using adaptive testing, whereby questions vary depending on a student’s prior answers, as part of this computer-based model.

3. In fact, we can look at the changes in those students' test scores and validly attribute the changes in those scores to specific people (teachers and principals) and institutions (schools) and again make high-stakes decisions based on those results.
4. By putting these test scores at the center of high-stakes accountability systems, we will create improvement incentives for students, teachers, principals, and others.
5. At the same time, we will not create incentives that undermine our overall educational goals, such as incentives to eliminate or reduce non-tested but still important content and types of learning.
6. The end result of all this will be improved academic outcomes.

Each of these six assumptions deserves some attention before this chapter shifts to an examination of school choice.

Scope of Test Coverage: Standardized testing in the past has overwhelmingly focused on mathematics and the language arts, excluding goals related to citizenship and the arts as well as the social sciences and, to a large extent, the sciences. These tests have also focused on readily tested knowledge and skills, to the exclusion of applied knowledge and deeper skills that are more likely to emerge and be used in project-based endeavors. A goal of the two new CCSS testing consortia is to assess a deeper and more applied set of knowledge and skills.

Precision, Accuracy, and Validity: States have used tests to retain students in grade, to place students in gifted programs, to grant admission to competitive schools, to assign students to high- or low-track classes, and to grant or withhold a high school diploma. For all of these uses, the assessments undoubtedly provide useful and relevant information. But the measurement error and the issues addressed above regarding test coverage raise serious issues about whether the assessments can be validly used for these purposes. The Joint Standards for testing, published by the American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (1999), caution that when test scores are used for such high-stakes purposes, “empirical evidence documenting the relationship among particular scores, the instructional programs, and desired student outcomes should be provided” (American Educational Research Association et al. 1999, p. 147). The Joint Standards also warn, “As the stakes of testing increase for individual students, the importance of considering additional evidence to document the validity of score interpretations and the fairness in testing increases accordingly” (American Educational Research Association et al. 1999, p. 141).

High-Stakes Attribution of Changes in Test Scores to Teachers and Principals: As noted above, students' test scores cannot be validly used for such purposes (Baker et al. 2010; Braun 2005; Corcoran 2010; Rothstein 2009). Classroom-based factors such as teacher quality likely account for no more than 20 % of the observed variance in students' test scores (see Hanushek et al. 1998; Rowan et al. 2002). Isolating that effect and then attributing it to the teacher—rather than to such features as peer effects, class size effects, or classroom resource effects—are beyond the capacity of any existing regression model. While current policies base high-stakes personnel evaluations on multiple criteria, using test-score growth for 20–50 % of

the overall score, the range of scores arising from the other criteria (e.g., classroom observations) tends to be less than the forced range of scores derived from the growth models, meaning that those models tend to have an outsized effect over final ratings.

Improvement Incentives: When NCLB was first being implemented, there were already concerns that it would impel schools to “teach to the test” (see discussion in Nichols and Berliner 2007). In response, the law’s defenders pointed out that prior to the law’s passage, many schools were almost completely ignoring the academic needs of their disadvantaged students. Teaching to the test would be a substantial improvement, particularly if states adopted tests worth teaching to. Supporters of test-based accountability systems also push back against the premise of this critique, arguing instead that schools will attempt to improve test scores by improving instruction, not by trying to “game” the tests.

Unintended Consequences and Counterproductive Incentives: According to Campbell’s Law, “The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor” (see Nichols and Berliner 2007). That is, as higher and higher stakes are attached to students’ test scores, two things can be expected to happen. First, the validity and usefulness of the scores themselves are undermined. Second, the teachers and others who are subject to the high stakes will change their behavior to respond to the new incentives, often in ways that are not beneficial to teaching and learning.

These two outcomes are interrelated. When teachers and principals are told that their success will be overwhelmingly judged by test scores, they will change classroom instruction and the allocation of learning resources in order to raise those scores. Some of these changes are straightforward: find out what will be tested and how it will be tested, and prepare students to master that material and those skills. Another change is completely unacceptable, no matter how predictable: cheating. But the most radical changes are deep and structural: an imbalance in curriculum and instruction. Curriculum is changed to add extreme focus on the tested content, primarily reading and mathematics. Instruction is similarly changed to add extreme focus to the type of knowledge and application demanded by the tests. Project-based, authentic learning is squeezed out, as is applied, deep knowledge.

Given such changes, what is then measured by the high-stakes tests? Students attending a school that resists these pressures and continues with balanced curriculum and instruction will be administered the same assessments as those attending schools with a laser-like focus on those tests. All things equal, the second school will show higher scores, but would lawmakers be correct in assuming that the second school is providing a better education? Or would much of the difference be best understood as a clear illustration of Campbell’s Law?

Schools and Educators Will Drive Better Outcomes: The pattern typically observed when high stakes are attached to a test is a quick improvement in those test scores followed by little or no subsequent improvement (Bowie and Green 2012). When academic performance is measured, however, on a different (low-stakes) assessment—in the USA, the relevant test fitting this description is the

NAEP—the early, quick improvement does not show up. It seems, then, that the early improvement is due much more to teaching students how to excel on a given high-stakes test than to any meaningful improvement in teaching and learning. The post-NCLB trends on the NAEP have been very similar to the pre-NCLB trends (overall, these are upward trends), suggesting that reform had very little effect on student learning (Fuller et al. 2007; but see Dee and Jacob 2011, suggesting some positive effects in math for some younger students).

3.2 *School Choice*

School choice comes in a wide variety of forms, including magnet schools, home-schooling, cyberschools, charter schools, open enrollment among schools within a public school district, inter-district public school choice, vouchers, and what I call “neovouchers”, which provide private school tuition assistance through a complex tax credit mechanism (Welner 2008). Almost all of these choice approaches are on the upswing, with neovouchers, cyberschooling, and charter schools showing particularly steep curves (Miron et al. 2012). Currently, more than 13 million US students exercise one or more types of school choice.

However, while school choice is discussed here as a “policy”, it is probably more accurately thought of as a “policy tool”. There are large differences between these different types of school choice. Further, for any given approach, there can be large and important differences depending on how the particular policy is structured (see Arsen et al. 2000; Welner 2008). When thought of as a policy tool, therefore, school choice can be crafted in ways to help accomplish the larger goals of the broader policy.

Yet the reality is that lawmakers in the USA have generally been pursuing school choice as an end in itself. The strongest choice advocates share much of the late economist Milton Friedman’s faith in the power of the free market to drive greater efficiency—as well as higher quality and even equity. Seven primary assumptions underlie this advocacy:

1. Increased choice will lead to innovations that are then used to improve the larger school system.
2. Increased choice will lead to competition that drives improvements in the larger school system.
3. Parents will choose schools that are better fits for their families and their children.
4. Parents will choose schools that increase their children’s academic achievement.
5. Increased choice will address inequity by providing new options for disadvantaged families.
6. Increased choice will not undermine our overall educational goals by, for example, increasing segregation and stratification.
7. In a choice system, market forces will hold schools accountable to families and will deter fraudulent or negligent behavior.

Briefly consider the evidence regarding each of these assumptions.

Choice Driving Innovation in Curriculum or Instruction: School choice has led to greater options for many parents. For example, models such as Montessori and Core Knowledge are often adopted by choice schools, giving more parents access to these schooling approaches. These are, however, not innovations of the sort imagined by early advocates of, e.g., charter school reform. The schools are adopting existing innovations. Professor Christopher Lubienski of the University of Illinois contends that the relative lack of innovation in the charter school sector is predictable, since each school's success depends on convincing parents to enroll their children (Lubienski 2012). A recognizable approach or model (Montessori and Core Knowledge being prime examples) will send a more attractive message to these parents: "We've come up with a great new idea, so please send your children to us so that we can try it out on them". That said, charter schools have been in the forefront of some true innovations, most notably the technology-based idea of "blended learning" (see Pandolfo 2012). Whether this new approach is beneficial is still an open question, but it is reasonably described as an innovation.

Competition Driving Improvement: Several studies have suggested small test-score benefits of competition in some states or cities. Others have found no effects of competition, even in areas with considerable school choice. Still others have found that competition has forced schools to devote limited resources to non instructional purposes such as advertising. And still others have raised concerns that competition for students does not play out equitably. That is, in a competitive system, some children cost more to educate or are less likely to have high test scores or other desired outcomes—many high-quality schools will therefore not compete for them. Others might, but the result would be an increasingly stratified system. Overall, the research results studying competitive effects of school choice "are mixed and inconclusive" (Arsen and Ni 2012, p. 194). Because the more aggressive forms of school choice are still relatively small in scope in most communities, it is difficult to say how many of these hopes and concerns will play out. For example, "while charter schools have garnered considerable attention over the past two decades, they only account for around 4.1 % of the national enrollment in public schools" (Miron and Welner 2012, p. 8).

Parental Choice of Fitting Schools: Research has consistently shown that parents are generally happy with the choices they make (see Mayer et al. 2002). It is also reasonable to assume that many choosers seek out schools with curriculum and instructional approaches that are good fits with their parenting philosophies.

Parental Choice of Academically High-Quality Schools: Parents are more likely to seek out schools with high test scores, but preferences and perceptions of quality appear to be influenced by race and other social demographics (Lacireno-Paquet 2012). That is, parents may view the race and poverty of children at a school as a proxy for quality. Even achievement test scores are a problematic proxy for quality of instruction, given that students' entering test scores—and their existing knowledge and skills—are a better predictor of subsequent test scores than any differences in

instructional quality between schools. In fact, there is little evidence to support even the basic assumption of a benefit to actively choosing a school other than one's assigned neighborhood school. The achievement results of school choice have been studied extensively for two prominent policies—charter schools and private school vouchers—and neither appears to be associated with appreciable improvements in average achievement outcomes (Miron and Urschel 2012).

Choice as Beneficial for Disadvantaged Families: As discussed above, some lawmakers in the USA have looked at equity from the perspective of the individual: do we have laws that treat all individuals equally? From this perspective, school choice can advance equity by removing some financial barriers for disadvantaged families. But if equity is framed as overall opportunities to learn, the picture is less rosy. As already noted, there is little evidence that exercising school choice provides academic benefits, so school reforms focused broadly on improving neighborhood public schools appear to be a wiser use of resources. Moreover, as discussed immediately below, school choice systems can exacerbate overall inequalities, segregation, and stratification.

Possible Unintended Consequences: School choice can take a variety of forms, and the specifics of any given policy can result in large differences in the policy's effects. School choice can, for example, be constrained so as to give priorities to low-income families or to families who live in given neighborhoods. Although unconstrained choice will likely lead to balkanized schools—with racial segregation as well as stratification by wealth, special needs status, and English learning status (Mickelson et al. 2012)—approaches that deliberately shape constraints can increase diversity in schools. More broadly stated, if school choice is used as a tool within a larger policy designed to accomplish broader societal or learning goals, the chances of unintended negative consequences can be greatly reduced.

Accountability Through Market Pressures: Ideally, a fully informed and efficacious set of school “consumers” will hold choice schools accountable by choosing only the highest-quality operators. In reality, many sketchy and low-quality operators have run choice schools and even thrived within the marketplace (see, e.g., Coutts 2011). Some of the worst offenders, who have committed fraud and other crimes, have eventually been shut down by regulators. But the overall marketplace reflects a wide distribution that includes a large number of high-quality, mediocre, and low-quality choice schools. This should not be surprising. People respond to incentives, but those incentives do not always play out as lawmakers might hope. Market incentives can, for instance, drive greater efficiency and even innovations. But the best efficiencies might come from screening out or pushing out students who are more challenging or expensive to educate, and innovations might be focused on lessening costs in ways that actually undermine learning goals. Market “accountability” can be effective, but so can bureaucratic accountability, accountability tied to performance goals, and accountability through professional standards of practice (Garn and Cobb 2012). Each approach comes with strengths and weaknesses.

4 Conclusion

The above summary of research, if written by a scholar more receptive to the current US reform agenda, would surely leave a somewhat more positive impression. Each researcher or author will bring his or her own perspective. But as long as we all stick to the overall body of research evidence, the differences in our summaries would only be at the margins. The truth would remain that the evidentiary record for test- and choice-focused policies gives us little reason to believe that either approach will yield substantial benefits. In fact, for those of us who are more skeptical, the evidentiary record provides clear warnings of unintended, negative consequences.

Yet these two reform locomotives show few signs of decelerating. Some backlash against test-based accountability is now developing (see Bui 2012), but even after the wave subsides, it is likely to leave behind considerable debris. Already, both reforms have reshaped the US educational system, and choice policies in particular become more pervasive and influential each year. When we look back after another decade, we will undoubtedly see considerable testing and choice; the open question is whether additional efforts—ideally focused on opportunities to learn—are on the ascendency.

References

- American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Arsen, D., & Ni, Y. (2012). The competitive effect of school choice policies on public school performance. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Arsen, D., Plank, D., & Sykes, G. (2000). *School choice policies in Michigan: The rules matter*. East Lansing: Michigan State University.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from http://epi.3cdn.net/724cd9a1eb91c40ff0_hwm6ij90.pdf
- Bowie, L., & Green, E. L. (2012, July 22). Baltimore school test results hit three-year lull: Results focus attention on Alonso's reforms. *The Baltimore Sun*. Retrieved from http://articles.baltimoresun.com/2012-07-22/news/bs-md-ci-school-progress-stalls-20120722_1_jessica-shiller-ceo-andr-s-alonso-maryland-school-assessments/2
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton: Educational Testing Service. *Brown v. Board of Education of Topeka, Kansas*, 347 US 483 (1954).
- Bui, L. (2012, November 15). Montgomery superintendent Joshua Starr's prominence injects him into national debate. *Washington Post*. Retrieved from http://articles.washingtonpost.com/2012-12-15/local/35847170_1_standardized-tests-new-national-curriculum-standards-teacher-performance
- Chen, K. (2012, November 30). Latest release of US High School graduation rates still preliminary. *PBS NewsHour*. Available online at <http://www.pbs.org/newshour/runtdown/2012/11/latest-release-of-high-school-graduation-rates-still-too-preliminary.html>

- Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value added measures of teacher effectiveness in policy and practice*. Annenberg Institute for School Reform. <http://annenberginstitute.org/pdf/valueaddedreport.pdf>
- Coutts, S. (2011, April 9). Akron's White hat gets poor grades for students. *ProPublica & Akron Beacon-Journal*. Retrieved from <http://www.ohio.com/blogs/education/first-bell-on-education-1.286050/akron-s-white-hat-gets-poor-grades-for-students-1.286681>
- Darling-Hammond, L. (2007). The flat earth and education: How America's commitment to equity will determine our future. *Educational Researcher*, 36(6), 318–334.
- Dee, T. S., & Jacob, B. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446.
- Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge no child left behind? *Educational Researcher*, 36(5), 268–278.
- Garn, G., & Cobb, C. (2012). School choice and accountability. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Hanushek, E. A., Kai, J. F., & Rivkin, S. J. (1998). *Teachers schools and academic achievement*. NBER working paper 6691. Cambridge, MA: National Bureau of Economic Research. From http://www.cgp.upenn.edu/pdf/Hanushek_NBER.PDF
- Lacireno-Paquet, N. (2012). Who chooses schools, and why? The characteristics and motivations of families who actively choose schools. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Lubienski, C. (2012). Educational innovation and diversification in school choice plans. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Mann, H. (Ed.). (1868). Twelfth annual report to the Massachusetts State Board of Education, 1848. In *Life and works of Horace Mann* (Vol. 3, p. 669). Boston: Walker, Fuller.
- Mathis, W. (2009). *NCLB's ultimate restructuring alternatives: Do they improve the quality of education?* Boulder: National Education Policy Center. Retrieved from <http://nepc.colorado.edu/publication/nclb-ultimate-restructuring>
- Mathis, W. (2012). *Research-based options for education policy making: Common core state standards*. Boulder: National Education Policy Center. Retrieved from <http://nepc.colorado.edu/publication/options>
- Mayer, D., Peterson, P., Myers, D., Tuttle, C. C., & Howell, W. (2002). *School choice in New York city after three years: An evaluation of the school choice scholarships program, final report*. New York: Mathematica Policy Research, Inc. Retrieved from <http://www.mathematica-mpr.com/publications/PDFs/nycfull.pdf>
- Mickelson, R. A., Bottia, M., & Southworth, S. (2012). School choice and segregation by race, ethnicity, class, and achievement. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Miron, G., & Urschel, J. (2012). The impact of school choice reforms on student achievement. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Miron, G., & Welner, K. G. (2012). Introduction. In G. Miron, K. G. Welner, P. Hinchey, & W. Mathis (Eds.), *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Miron, G., Welner, K. G., Hinchey, P., & Mathis, W. (Eds.). (2012). *Exploring the school choice universe: Evidence and recommendations*. Charlotte: Information Age Publishing.
- Moreno, A. (2012). *Does retention (repeating a grade) help struggling learners?* Denver: Marisico Institute for Early Learning and Literacy. Retrieved from http://www.du.edu/marsicoinstitute/policy/Does_Retention_Help_Struggling_Learners_No.pdf

- National Center for Educational Statistics (n.d., a). *Fast facts*. Available online at <http://nces.ed.gov/fastfacts/display.asp?id=65>
- National Center for Educational Statistics (n.d., b). *Ed Data express: National snapshot*. Available online at <http://www.eddataexpress.ed.gov/state-report.cfm?state=US&submit.x=21&submit.y=8>
- National Center for Educational Statistics (n.d., c). *Fast facts*. Available online at <http://nces.ed.gov/fastfacts/display.asp?id=372>
- National Center for Educational Statistics (n.d., d). *The condition of education: Education expenditures by country*. Available online at http://nces.ed.gov/programs/coe/indicator_ifn.asp
- Nichols, S., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge: Harvard Education Press.
- Oakes, J., Quartz, K. H., Ryan, S., & Lipton, M. (2000). *Becoming good American schools: The struggle for civic virtue in school reform*. San Francisco: Jossey-Bass.
- Pandolfo, N. (2012, September 24). Education Nation: In Arizona desert, a charter school competes. *Hechinger Report*. Retrieved from http://hechingerreport.org/content/education-nation-in-arizona-desert-a-charter-school-competes_9687/
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537–571.
- Rowan, B., Correnti, R., & Miller, R. J. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools*. CRPE Research Report RR-051. From <http://cw.marianuniversity.edu/mreardon/755/document%20repository/Teacher%20Effects%20on%20Student%20Achievement.pdf>
- Smith, M., & O'Day, J. (1990). Systemic school reform. In *Politics of education association yearbook* (pp. 233–267). Taylor & Francis: London.
- Tyack, D. (1974). *The one best system*. Cambridge: Harvard University Press.
- Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(1), 1–19.
- Wells, A. S., & Serna, I. (1996). The politics of culture: Understanding local political resistance to detracking in racially mixed schools. *Harvard Educational Review*, 66, 93–118.
- Welner, K. G. (2001). *Legal rights, local wrongs: When community control collides with educational equity*. Albany: SUNY Press.
- Welner, K. G. (2008). *NeoVouchers: The emergence of tuition tax credits for private schooling*. New York: Rowman & Littlefield.
- Welner, K. G. (2013). Consequential validity and the transformation of tests from measurement tools to policy tools. *Teachers College Record*, 115(9). http://nepc.colorado.edu/files/welner_tcr_consequential_validity.pdf
- Welner, K. G., & Oakes, J. (2007). Structuring curriculum: Technical, normative, and political considerations. In F. Michael Connelly, M. F. He, & J. Phillion (Eds.), *Handbook of curriculum and instruction* (pp. 91–112). Thousand Oaks: Sage. Chapter 5.
- Welner, K. G., Biegel, S., & Jackson, D. D. (2010). Legal foundations of education. In S. Tozer et al. (Eds.), *Handbook of research on the social foundations of education* (pp. 153–162). New York: Routledge/Taylor & Francis.
- Yettick, H., Love, E. W., & Anderson, S. (2008). Ineffective decision making and educational opportunity. In K. Welner & W. Chi (Eds.), *Current issues in education policy and the law* (pp. 99–120). Greenwich: Information Age Publishing.