

CHAPTER 1

A SHORT HISTORY OF HIGH-STAKES TESTING, Arguments For and Against Its Use, Its Place in Contemporary Society, and a Brief Introduction to Campbell's Law

The commission investigating the events leading up to the tragedy of September 11, 2001, found evidence that warnings of the impending violence were clear and numerous. These warnings were unheeded by federal agencies and most of the people who staff them because no crisis seemed imminent. Even then, however, amid the relative peace, some individuals tried to alert the nation that we should not be complacent. These individuals warned us that a crisis was indeed brewing and that the United States would surely be tested.¹

Once again the United States faces a severe crisis, once again the dangers do not seem imminent, and once again politicians and federal agencies are being alerted to the danger. This time it is not merely a few individuals but literally thousands, particularly teachers, who believe there is danger. Sadly, as in the past, the counsel of these many individuals is unheeded by those who wield political power.

This crisis concerns the corruption of what is arguably America's greatest invention—its public schools. This book joins with others in documenting the damage to education caused by overreliance on high-stakes testing.² Our research suggests that the incidence of negative events associated with high-stakes testing is so great that, were we the Centers for Disease Control, we would be forced to declare an epidemic. As will be made clear, because of high-stakes testing, public education is presenting serious and harmful symptoms. Unlike other critics of the high-stakes

testing movement, however, we demonstrate that a powerful social science law explains the etiology of the problems we document. Ignorance of this law endangers the health of our schools and erodes the commitment of those who work in them.

A SHORT HISTORY OF LOW-STAKES AND HIGH-STAKES TESTING

For about a century, standardized testing for assessing aptitudes (e.g., intelligence) and achievement (e.g., mathematics) has played an increasingly prominent role in shaping American educational thinking.³ Today such tests are common. Among the well-known tests used throughout America's primary and secondary grades are the Iowa Test of Basic Skills (ITBS), TerraNova, Stanford Achievement Tests (SAT), and the Metropolitan Achievement Test (MAT). In addition, generations of secondary students have prepared for college entrance by taking the SAT exam administered by the College Board or the ACT exam administered by the American College Testing Program. Applicants applying for entrance to graduate school are often required to take other exams, usually the Graduate Record Examination (GRE) and/or the graduate tests for law (LSAT), business (GMAT), or medicine (MCAT).

These various tests have been used to assess students' aptitudes and achievements, inform decisions about curriculum and instruction, and make predictions about how successful a student may be in the future. Most Americans have had faith that these tests were crafted with the utmost care, with proper concern for validity (does the test really measure what it purports to measure?) and reliability (are the scores students get dependable, were the test to be given again?). For the most part, these tests have been successful. Most citizens are satisfied with the information received from these tests and the uses that are made of them.

Except for the graduate school entrance examinations, these various tests were not often high stakes. That is, they rarely had dramatic and life-changing consequences attached to the scores that were obtained from them. For decades children were not left back a grade or denied a high

school degree solely on the basis of their test scores, though that did occasionally happen. On the basis of standardized tests of achievement, however, students were sometimes denied or admitted to gifted programs, and the tests often determined whether or not remedial programs were appropriate for a student. But for most students and their families, achievement tests provided information, not consequences. In recent decades teachers and administrators were rarely given bonuses or fired for the performance of their students on standardized tests, though this too occasionally did happen.

It is true that college entrance examination scores sometimes prevented a student from attending the college of his or her first choice. But rarely, if ever, was a student prevented from going to a public community or public four-year college solely on the basis of his or her score on a college entrance examination. Over the last half-century, regardless of one's test score, opportunities for obtaining a college degree existed. For most Americans, there were very few high-stakes decisions made about young students, their teachers, schools, and curricula on the basis of achievement test scores alone. The times have changed.

The expansion of high-stakes testing

The current emphasis on using tests for making important decisions about students, teachers, and administrators in the elementary and secondary schools—and also for evaluating the schools and school systems those students attend—can be traced back to the 1965 authorization of the Elementary and Secondary Education Act (ESEA).⁴ In the post-*Sputnik* era, ESEA was the government's answer to the call for greater attention to the quality of America's schools and the needs of students from less advantaged homes. The concerns emanating from Russia's win in the space race resulted in the development and implementation of minimum competency tests—tests used to ensure all students left school with at least the ability to read and do basic math. Students could be denied a diploma if they did not pass these tests, but there were no consequences for teachers or schools. Eventually the minimum competency tests were criticized for

being relatively easy to pass since they were concerned with minimums to be learned: the achievement floor and not the achievement ceiling. Many politicians and citizens alike believed that students were not being stretched enough, that overall U.S. school performance was not improving, and that the terrible achievement gap between white middle-class students and black, Hispanic, or poor students was not being reduced. As discontent grew, the language of minimum competency changed to the language of standards-based achievement.⁵

In the years following ESEA, concern for American education grew. This was fueled, in part, by international data that purported to show that our schools were not as good as those in other nations. More fuel was provided because the international economy was growing at a time when the national economy was performing poorly. A scapegoat for that state of affairs needed to be found. The concerns of the 1970s culminated in the 1983 release of *A Nation at Risk*, a report that predicted that unless public education received a major overhaul and unless expectations for student achievement were raised, America's economic security would be severely compromised.⁶ American education became the scapegoat for a host of bad business decisions. As intended, that document sparked a renewed interest in and urgency about how we educate America's students. That sense of urgency set in motion a series of policy initiatives aimed at improving the American education system.

In the years following *A Nation at Risk*, many have eloquently and vehemently challenged the logical and empirical validity of the assertion that American education is badly or fatally flawed. In fact, hysteria about the achievements of our schools was, and continues to be, largely a myth.⁷ But the myth lives on, and policies follow from myth as surely as from factual accounts about the way the world works. Despite its mistaken factual claims, after publication of *A Nation at Risk*, many politicians aligned with a growing public demand to improve the "failing" educational system. As a result, the past 20 years have seen a broad range of policy documents and initiatives offering ways of solving America's educational problems; among these was a call for more consequential testing.⁸ As we write in mid-2006, we find that the American educational system is being completely transformed from what it was at the time of *A Nation at Risk*.

It is now clear that high-stakes testing holds a prominent place in that transformation, which is the reason for this book.

Some problems with high-stakes testing

The practice of high-stakes testing is not new, having been used in ancient China to hire civil servants after applicants for these jobs had invested years studying for the examination. As one might expect, then, criticisms of high-stakes testing are also not new. For example, the New York State Department of Education made the following comments to the state's legislature as the latter body contemplated establishing what we would today call a high-stakes test:

It is an evil for a well-taught and well-trained student to fail in an examination.

It is an evil for an unqualified student, through some inefficiency of the test, to obtain credit in an examination.

It is a great and more serious evil, by too frequent and too numerous examinations, so to magnify their importance that students come to regard them not as a means in education but as the final purpose, the ultimate goal.

It is a very great and more serious evil to sacrifice systematic instruction and a comprehensive view of the subject for the scrappy and unrelated knowledge gained by students who are persistently drilled in the mere answering of questions issued by the Education Department or other governing bodies.⁹

In somewhat archaic language, the Department of Education raises concerns about the reliability and validity of its tests, as every testing agency should. But they also are concerned about overtesting and about how testing programs can mislead students (and by implication parents and politicians) into thinking test scores are indicators of sound instructional practice. This Department of Education also expresses its worry about how testing can distort the basic purposes of education by narrowing the curriculum. The enlightened bureaucrats who wrote this report to

the legislature were warning the state's politicians that it is possible, even with the best of intentions, for testing programs to corrupt the educational process. The archaic language in this report is better understood after we discover when it was written: 1906!

Another warning about the dangerous side effects of high-stakes testing surfaced when a plan to pay teachers on the basis of their students' scores was offered, making student test scores very high stakes for teachers. A schoolmaster noted that under these conditions, "a teacher knows that his whole professional status depends on the results he produces and he is really turned into a machine for producing these results; that is, I think, unaccompanied by any substantial gain to the whole cause of education."¹⁰ This concern about testing students to judge a teacher's worth first surfaced in the year 1887, but it is as fresh as recent headlines about pay-for-performance in Denver, Colorado; Houston, Texas; Florida; Minnesota; and Iowa.¹¹

These worries about what we now call high-stakes testing were made well before modern testing systems had gained admiration for their alleged beneficial effects on education. No crisis seemed imminent, and so the minor worries about testing voiced by an individual or an organization here and there, over the last century, were easily set aside. But high-stakes testing in the United States is now more widespread than ever before in our history, and our nation apparently relies more on ability and achievement testing than most other nations for making important decisions about individuals and schools. Clearly, we live in an era and in a nation where there is strong support for public policies that use test results to compel changes in the behavior of students, teachers, and school administrators. Our president, politicians from both parties, and many citizens, particularly members of the business community, believe that education can be best improved by attaching consequences (i.e., attaching high stakes) to tests. The tests are seen by some as the perfect policy mechanism because they are both *effectors* and *detectors*—they are intended to *effect* or cause change in the system and then *detect* whether changes in the system actually occur. The federal No Child Left Behind Act of 2001 (NCLB) sanctified this approach to school change for every state, district, and school.¹²

As we demonstrate throughout this book, the logic undergirding high-stakes testing policy is unsound. We will show by example after example why strict adherence to this policy does more harm than good. Guided by a little-known but well-documented principle of social science, Campbell's law, we will argue that high-stakes testing, the cornerstone of NCLB, is paving the way for an educational crisis that threatens to leave our *nation* behind.

The present status of NCLB and high-stakes testing

The No Child Left Behind Act, passed in 2001 and signed into law in January 2002, is the reason for the present spread of high-stakes testing. This law is probably the most invasive and complex piece of federal legislation on education in our nation's history. As it has developed and influenced our educational system, we and others have argued that NCLB

- is flawed legislation and cannot work as designed;¹³
- does not produce gains on the National Assessment of Educational Progress (NAEP tests) or other tests of achievement that can be used to audit a state's performance;¹⁴
- is not closing the achievement gap as it was intended to do;¹⁵
and
- is increasing dropout rates among our nation's youth.¹⁶

The faults of the NCLB legislation are numerous, but they are *not* the major concern of this book. We are concerned here with only one of the provisions of NCLB, the one requiring that states adopt a system of accountability whereby students, teachers, administrators, and schools are evaluated annually on the basis of students' standardized test performance and that consequences follow when student scores are low or annual gains in school achievement are not made.

Stakes are "high" because of the life-changing significance of the consequences attached to test scores. For example, the consequences of low scores for students include failure to be promoted to a subsequent grade, failure to graduate high school, or denial of college scholarship monies.

Low-scoring students can also switch schools, receive tutoring at school expense, and may also have to attend district-mandated Saturday school, summer school, or after-school programs. Teachers and administrators can receive bonuses for high student scores, or, as is more typically the case, they can be reassigned or fired because of low student scores and poor student gains. Low scores and poor gains allow for schools to be reconstituted as public charter or private schools or simply closed, with the students reassigned elsewhere. The public ratings of public schools also mean that all school personnel receive public accolades or a public scolding as a consequence of school and district test scores. Shaming, an ancient ritual, is a component of NCLB.

Scholars and politicians from divergent viewpoints generally believe that NCLB is legislation badly in need of serious change.¹⁷ As noted, however, our concern is with just one piece of the NCLB legislation required of all the states, sometimes even eagerly accepted by them, that may live on after the inevitable restructuring (or demise) of NCLB. Our concern is with the acceptance of high-stakes testing as the mechanism to effect changes in our schools. In the chapters that follow, we present evidence that high-stakes testing so distorts and corrupts education that their continued use seriously endangers the educational profession and limits the learning outcomes of our youth.

THE LOGIC, ILLOGIC, AND PARTIAL LOGIC USED IN DEFENSE OF HIGH-STAKES TESTING

The theory of action embedded in NCLB is that a system of threats and incentives tied to test performance will energize teachers and their students into working harder and more effectively.¹⁸ Threats and incentives of this type were commonly used to manage workers in the nineteenth century—a workforce consisting mostly of laborers whose intellects were not valuable and who were completely replaceable (pieceworkers in shirt factories, fruit pickers, nonunion laborers). But modern businesses have abandoned this approach as society came to rely more on the intellectual capital of its workers. Computer designers, economists, and market-trend

watchers are rarely managed with threats of punishment. Although rewards/bonuses are sometimes available, rarely are threats relied on as a way to spur these kinds of workers into action. After all, businesses don't want to drive out a labor force that requires years of training and for whom productivity cannot be defined by a single indicator.

The theory of action behind the modern corporation dealing in knowledge and creativity on a global scale is quite different from the one designed to motivate people doing work that is inherently uninteresting and for which replacements are easy to obtain. So incentives and threats seem an inappropriate theory to use with teachers and their students, unless they are considered to be more like laborers than they are like knowledge workers. That distinction seems lost on proponents of this archaic theory of management who argue that carrots and sticks will address problems of low expectations for poor and minority students, will minimize the unequal educational opportunities between the haves and have-nots, and will help establish a clear set of state standards to guide curriculum and instructional decisions. For policymakers and citizens yearning for simple solutions to education's complex problems, a management strategy characterized by reward and punishment is easily embraced.

Let us look a little closer at ten of the many arguments offered by supporters of the NCLB legislation.¹⁹ These arguments are derived from the flawed theory of action that undergirds the legislation.

1. *Students work harder and learn more when they have to take high-stakes tests.* Response: This is true for some students but certainly not for all. Our experience in low-income and high-income schools alike leads us to believe many students don't work harder at all, finding test preparation boring. Intense and lengthy preparation for high-stakes testing reduces most students' motivation for learning and thus reduces their commitment to do the work necessary to score higher on the test. And in many schools, irrespective of the social class makeup of the students, it is not clear that more is learned through implementation of high-stakes tests. A good case can be made that state standards-based curricula (required by NCLB) narrow the curriculum to the subjects in which students will be tested. That is, if graphing is not on the state mathematics test, there seems to be no need to teach anything about graphing. Students too often seem

to learn what is on the high-stakes test, and no more. In addition, it is quite clear that the high-stakes testing movement has reduced the overall curricula choices of students. Where once we saw school time devoted to music, art, physical education, and social studies, now we see a vast reduction in these offerings as more time is spent on the areas of the curriculum that are tested and therefore consequential.²⁰ So it is not at all clear that students learn more when high-stakes testing is implemented or that they are more motivated to try.

2. *Students will be motivated to do their best and score well on high-stakes tests.* Response: The current version of NCLB does not reward excellent performance on high-stakes tests. Instead, it rewards performance above a certain passing score called "proficient." Once that level of proficiency is set, that is all that counts for federal purposes. So to avoid sanctions by the federal government, many states have apparently set the "proficient" level on their state tests at a relatively low level of achievement.²¹ In those states where the standards are low, therefore, there is scant motivation for students—or for teachers to help students—to do their best. There is simply no incentive to go beyond what is minimally necessary to score at the "proficient" level on the test. The testing program we have bought into destroys motivation to excel, rewarding, instead, motivation to get by. Furthermore, our talks with students reveal that many are angry they will be tested by the state and, on the basis of their test score, could be held back or denied a high school degree. They say that if they have done their homework, taken and passed class tests, not been absent or a troublemaker, they should be passed or graduated. They say that their teacher knows what they can and cannot do and that a test imposed from outside is unfair. Although certainly not in their own self-interest, this student anger about the tests often seems to result in lowering motivation to do well on the tests instead of increasing motivation to do well.

3. *Scoring well on high-stakes tests leads to feelings of success by students, while doing poorly on such tests leads to increased effort to learn.* Response: In an ideal world, students attribute their successes on tests and in life to the hard work they did, and they also attribute their failures on tests and in life to their own lack of preparation and laziness. In this ideal world, students would take pride in success on a test and would commit themselves