

ways our schools operate. So scores on tests will mislead us about genuine productivity. Nevertheless, it all sounds sensible and thus appeals to many citizens who end up supporting the use of high-stakes testing programs for our schools.

A second and related reason high-stakes testing has slipped into mainstream culture is the emerging belief by both business and government that the future economy depends on a highly educated workforce. This belief took on new urgency after Thomas L. Friedman's book *The World Is Flat* became a best-seller.²⁷ Large numbers of the citizenry now believe that we need to push all our children to the farthest heights of education, moving most children to high school graduation with a degree indicating that a rigorous curriculum had been mastered. After that, most of those graduates need to be moved into degree-granting two- and four-year colleges. Obviously, the demand for a rigorous curriculum and college-level preparation means a seriousness about testing in our public schools never before required. High-stakes testing is compatible with these national ambitions. High-stakes testing fits neatly into the mindset that Americans have—namely, that to be competitive in the interdependent global economy we need high rates of college graduation, especially in the fields of science and engineering. In fact, this whole theory may be wrongheaded. As Denis Redovich has reported in article after article, the employment profile of the future does not support the need for a big increase in the mathematical and scientific knowledge of our youth.²⁸ We may well be overdemanding in these areas and already producing enough scientists and college graduates for the needs of the economy. A scientifically sophisticated citizenry is certainly in our national interest, but making advanced mathematics and science a major goal of U.S. education may be counterproductive, producing large increases in school dropouts and contributing to student anomie. In fact, college graduates have recently suffered from the same drop in wages that first hit those who failed to graduate from high school decades ago. That drop was followed only a short time later by a drop in the wages for those who had only a high school education. Now, those with college degrees are suffering the same fate. Earnings for workers with four-year degrees fell 5.2 percent from 2000 to 2004 when adjusted for inflation, according to White House economists.²⁹ Apparently, large percentages of

recent college graduates are taking jobs for which no college degrees are necessary, and the trend may be accelerating. Nevertheless, we continue to demand that the educational system produce ever-increasing rates of high school and college graduates, though we may actually now be near record levels of high school graduation rates.³⁰

Despite the fact that our national productivity is much more dependent on our tax structures, relative lack of corruption, and remarkable entrepreneurship, the citizenry believes that we need better schools to be competitive in the world economy. High-stakes tests therefore are a policy mechanism attuned to the mindset of many of our citizens in power and thus slip easily into our everyday life.

The third reason for the ease with which high-stakes tests have become commonplace in our culture is because of changing demographics. We can now clearly see the shape of an emerging gerontocracy. An older citizenry, much whiter than the youth of the nation, relatively well-off financially now but likely to live well beyond their resources, is beginning to act politically in their own best interests.³¹ As a powerful political and economic force, these folks want income and services. They will demand medical, pharmaceutical, and social services, full payment of social security, and some form of housing support as their income stays relatively fixed. They may not want to spend, nor will there be much money to spend, on lazy and unappreciative youth, especially youth of color, the offspring of other peoples. For many advantaged people, high-stakes testing separates the deserving poor from the undeserving poor and becomes a policy mechanism to preserve social status more than a policy to improve our schools. High-stakes testing subtly fits the mindset of this growing demographic and thus makes it easier for this policy to gain purchase in our contemporary society.

A fourth reason is based on the same self-preserving, privilege-maintaining quest. Simply put, the power elite in this society, along with the vast middle and upper-middle class whose children now attend good public schools, see high-stakes testing working to their own children's advantage. While they bristle that their own children must suffer through these tests (witness the Scarsdale, New York, mothers' rebellion),³² the schools their children attend are not much bothered by the tests and

the pass rates for their children are very high. When their children are in danger of not succeeding on such tests, they have the intellectual resources to help and can also pay for the tutors needed to ensure their children's success. High-stakes testing blends with their self-interest because it forces a kind of education on the children of the poor that ensures they cannot compete with the children of the wealthy. The drill and test-prep education we see in schools for the poor—and their high failure rate—does not prepare them for knowledge workers' jobs nor for the best universities in the nation. This makes room for the children of the more privileged. Since the status of children from advantaged families is preserved through high-stakes testing, it is easy for these folks to defend their use.

Fifth is the fit between high-stakes testing and other spectacles enjoyed by the public, such as baseball, football, basketball, and hockey. We are a game-playing, competition-seeking nation, and high-stakes testing fits easily into such a culture.

As is true of many sports, high-stakes testing has a tournament-like quality to it, thus bringing seasonal excitement to fans who now can follow the heavily publicized "winning" and "losing" streaks of their local schools, as they have often followed their local teams. Every summer when spring test results are released, there is World Series-like publicity and fanfare around how well (or poorly) our nation's teachers and students did that previous year. And like fanatic fans who delight in watching rivals have a losing season, American media feeds on whatever bad news exists,³³ prompting those who follow the news to ponder endlessly, Why have certain schools/teams failed? How many times has this school/team failed in the past? What is that school's/team's track record? What schools/teams might need to be reconstituted, closed down, or moved? What will we do to get rid of the bad teachers/players, and precisely which ones are they? Is it the science teacher or the first baseman, the English teachers or the defensive line, the coach or the principal? On whom, exactly, can we pin this failure?

Numerous similarities explain the country's preoccupation with testing. After all, the sport of cricket is called a test. Professional athletes in cricket and other sports practice hours and hours repeating the same activities endlessly so that their responses at "test" time will be automatic. In the high-stakes testing game, teachers also engage their students in

endless repetitive activities to better ensure students' responses are accurate and automatic come test time. In professional sports, teams with the highest-paid athletes are more likely to have winning seasons. Similarly, schools with more resources and those that serve the most affluent students tend to perform better academically.³⁴ (Perhaps that is why such inequalities in education are so easily accepted.) In professional sports, fans are immersed in statistics that highlight the successes and failures of their favorite teams and players (e.g., RBIs, shots blocked, touchdowns, assists). In the testing game, parents, politicians, and other community members are immersed in media coverage of academic data showing who is winning and who is losing.

Of course, we know stats say little about a player's many other contributions to the team, such as level of dedication, commitment, morale, and leadership. Similarly, when teachers and administrators are judged by their students' scores, often absent are a teacher's many other contributions such as their nurturance of a love for learning, individual attention given to counseling in a student's time of need, extra time spent meeting with students' families, provision of money from their own pockets for classroom items, and so forth.

High-stakes testing is now a part of our culture and has come to prominence, we think, because it fits easily into contemporary ways of thinking about our nation and ourselves. We are a political and an economic system dominated by the interests of big business, and so business models of accountability for our schools naturally follow. High-stakes testing seems to be a hardheaded business practice brought to bear on the schools, despite the fact that no one uses such a system in knowledge-oriented businesses. And unless we are greatly mistaken, schools fall into that category.

High-stakes testing also seems to help in preparing us for the vicissitudes of a competitive world economy, and so it is easily embraced. The fact that the new American economy may be vastly more service oriented than previously believed—and that it may not require nearly as many college graduates as is now thought necessary—is a point of view that is ignored.

The needs of the emerging gerontocracy and those who already have some status in society are also served by high-stakes testing. And high-stakes testing fits neatly into the gaming and spectacle-seeking culture

that so permeates U.S. society. For all these reasons, high-stakes testing has grown to be an acceptable part of the culture. Those who oppose the spread of high-stakes testing are seen as status-quo oriented, against quality in education, against school improvement, obstructionist, anti-efficiency, anti-George W. Bush, and so forth. But we are actually against high-stakes testing for none of these reasons. We are against it because teachers, students, parents, and American education are being hurt by high-stakes testing, and we have, in what follows, the evidence to demonstrate it. We are not against accountability or assessment. We are against accountability systems in education that rely upon high-stakes assessments.

The pages that follow are filled with examples of humiliation, devaluation, deceit, and other illustrations of the tragedies that accompany high-stakes testing. We foreshadow these stories and research reports by sharing a personal story (box 1.3), brought to one of the authors by a real teacher whose livelihood and dignity were threatened as a result of high-stakes testing. This tale illustrates the problem of judging teaching primarily on the basis of test results. Unlike athletes, who are judged mostly on the basis of their own performance, teachers are often held solely responsible for their *students'* performance.

What is problematic about this scenario and thousands of others like this one is not that the principal is monitoring teacher effectiveness. Nor is it problematic to want all teachers to do well. The problem is that these kinds of decisions about the quality of instruction and the success of a school are driven solely by student performance on a standardized test. Just like an athlete who contributes so much more to the team than just his/her statistics, a teacher contributes a great deal more to a student's educational experience, growth, and development than just what is reflected by a test score alone.³⁵

CAMPBELL'S LAW AND THE UNCERTAINTY ASSOCIATED WITH TESTS AND TESTING

As the examples presented earlier illustrate, and as common sense informs us, the pressure to do well on a single indicator of performance from which

Box 1.3 A True Story of Problems That Arise from High-Stakes Testing

A dedicated eighth-grade math teacher we know told us that in one year he went from being a celebrated, successful teacher, to being required to attend “remedial” teaching workshops. We asked, “What happened?” In the first year, he said, he taught students who were relatively motivated and interested in the subject. Although these students struggled throughout the year to grasp the mathematics he was teaching, their motivation and his teaching efforts resulted in significant learning, as reflected in his students’ “acceptable” test score performance. The teacher was asked to lead workshops to share his techniques with less successful colleagues.

The very next year, however, he saw an influx of students with special learning needs or for whom English was a second language. Still, he went to work doing everything he knew how to do—employing the same tactics that made him a “success” the previous year. He made more home visits than he ever had before and stayed after school to tutor as many students as possible—all without extra support. In the end, all that mattered were the test scores. The principal, seeing practice test scores that were consistently low throughout the fall and early spring terms, actually asked the teacher to attend the same workshops he once taught so he could “improve” his teaching. “How can we recognize good teaching and work to improve it,” the teacher asked us, “in an atmosphere of such confusion?”

This story was generously shared with us by a middle school teacher from San Antonio, Texas, who wishes to remain anonymous.

important consequences are derived can be especially counterproductive and destructive. But many of the problems encountered in situations like this are merely illustrations of a well-known social-science law directly applicable to the high-stakes testing associated with NCLB. We refer to this as Campbell’s law, named for the well-respected social psychologist, evaluator, methodologist, and philosopher of science Donald Campbell, who brought it to the attention of social scientists in 1975. Campbell’s law stipulates that “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 was intended to monitor."³⁶ Campbell warned us of the inevitable problems associated with undue weight and emphasis on a single indicator for monitoring complex social phenomena. In effect, he warned us about the high-stakes testing program that is part and parcel of NCLB.

George Madaus and Marguerite Clarke believed that Campbell had uncovered a kind of social-science equivalent of the Heisenberg uncertainty principle.³⁷ Named after its discoverer, Werner Heisenberg, the Heisenberg uncertainty principle states that there is always a degree of uncertainty associated with simultaneously measuring the position and velocity of microscopic objects. "The more precisely the position is determined," Heisenberg explains, "the less precisely the momentum is known in this instant, and vice versa."³⁸ To try to measure a microscopic object inevitably alters the conditions so much as to render the measurement inaccurate—there is inherent uncertainty in the data that is collected.³⁹

Madaus and Clarke noted that if Campbell is right, whenever you have high stakes attached to an indicator, such as test scores, you have a measurement system that has been corrupted, rendering the measurement less accurate. Apparently, you can have (a) higher stakes and less certainty about the validity of the assessment or (b) lower stakes and greater certainty about validity. But you are not likely to have both high stakes and high validity. Uncertainty about the meaning of test scores increases as the stakes attached to them become more severe. The higher the stakes, the more likely it is that the construct being measured has been changed. Applied to NCLB, scores on the instruments used as measures of academic achievement become less interpretable as the consequences for not making adequate yearly progress increase.

The ubiquity of Campbell's law

Evidence of Campbell's law is everywhere. Wherever we look, when high stakes are attached to indicators, what follows is the corruption and distortion of the indicators and the people who employ them. In business, economists have long recognized the possibility for corruption when stakes

are high. For example, incentives such as big bonuses for increased sales are common in business, but when stakes are attached to sales, then the business of selling could become corrupt. Cars may be sold that are lemons, houses may be sold with concealed defects, guarantees may be made that are not genuine. It is the sale that is important. In California, Sears once stopped paying salaries to its auto mechanics, instead putting them on commission for the repairs they sold to customers. It wasn't long before California authorities had to threaten to close Sears's auto shops because of all the complaints about unnecessary repairs that were being sold.⁴⁰ Other businesses have similar problems. Enron, Tyco, and Qwest, to name just a few examples, all needed to keep their stock prices high (their indicator of successful corporate management), and all did so by cheating. Shell needed its stock price high, so it overstated oil reserves. Pharmaceutical companies needed their share prices high, so they lied about the effectiveness of some of their drugs, and people died as a result.

When Russian factory workers in the old Soviet Union were ordered to produce so many tractors or tanks a month or face serious consequences, the requisite number of tractors or tanks were always produced. The fact that the quality of these machines was awful was irrelevant. Many simply did not work. But the quota had been reached and heads did not roll! That was what became most important to the factory workers.

When police in one city were ordered to clear up the number of unsolved crimes, the rate of solved crimes suddenly went up dramatically. The police simply made deals with newly captured criminals to confess to many crimes they did not commit. For their confessions, the criminals got lighter sentences. Crimes appeared to be solved, politicians were happier, but in fact the public was being fooled—it was all being done by smoke and mirrors.

In baseball, the Olympics, the Tour de France, and in many other sports, steroids have been banned, but they are still used repeatedly. Why? Because the indicators of success have great value attached to them. Monetary rewards and fame go to those that hit home runs, speed through water or over land, and can endure the rigors of many days of bicycling. Losers get little. So some athletes risk their health and their reputations to

make themselves look good on the indicators society values. As those indicators take on importance, Campbell's law seems always to come into effect, and corruption of the indicators of success and the athletes themselves follows. Some of the corruption of athletes starts early, as when they are recruited to colleges by being offered sex, parties, and monetary incentives for signing. Then they get all kinds of special privileges after arriving on campus, including high grades in courses specially designed for them. The importance of the winning column corrupts coaches, athletes, and the entire academic enterprise.

Citizens hope that their physicians, like their teachers, will be less corruptible than athletes and businesspeople. We depend on them for so much and generally hold them in high regard. Nevertheless, in medicine, when physicians are rewarded for the number of patients they see, rather than by their patients' perceptions of caring, there is an inevitable corruption of the medical care they give. Value has been attached to the wrong indicator, and thus physicians' caring behavior for their patients inevitably is compromised. Physicians face another problem with another indicator. Physicians are regularly evaluated on the two types of errors that they make. They can diagnose a healthy person as being sick (Type I error) or they can diagnose a sick person as being well (Type II error). Obviously, there are serious consequences associated with either type of error, but the consequences for making a Type II error are much more severe. Type II errors can result in expensive litigation and the loss of reputation by a physician or a hospital. Since the numbers of diagnoses of genuinely sick people as being well must be minimized by physicians, it should not be surprising to find out that physicians have found ways to avoid making Type II errors. So much weight is on that single indicator that physicians have found ways to rarely get caught making mistaken diagnoses.

Physicians avoid Type II errors by overdiagnosing illness among their healthy patients. That is, physicians purposefully inflate their Type I errors to deflate their Type II errors! The result of this is that many people believe they are sick when they are not. Treatments such as invasive surgery and pharmaceuticals are prescribed though they may not be needed.⁴¹ In addition, to keep Type II errors down, physicians also order tests for illnesses that are not likely to be found. Of course, the cost of medicine is

driven up enormously by these methods designed to avoid the dreaded Type II error. As Campbell's law predicts, the more importance that an indicator takes on, the more likely *it*, and the people who depend on it, will be corrupted.

We could cite hundreds more cases where Campbell's law seems to explain what is happening. There are military recruiters who must make quotas, and so they lie to young men and woman and their parents. There are scientists who are judged by the amount of grants they get, and so they fudge their data to get those grants. Government officials and businesspeople are sometimes judged by their university degrees, not their competence, so they take degrees from diploma mills. There are corporations that are judged by their safety records, and so many accidents go unreported, endangering the lives of their workers. There are child-protection agency workers who are judged by how many homes they visit per week. When resources are spread thin and the number of visits is given priority over meticulous analyses of a child's environment, children occasionally die.

Despite the sheer number of examples showing negative effects when particular indicators take on too much importance, many people still believe high-stakes testing is a viable way to improve education. By doing so, however, they defy a perfectly valid and ubiquitous social-science principle—at their peril. High-stakes testing is exactly the kind of practice that Donald Campbell warned us about. After all, high-stakes testing places significant value on test scores. Serious, life-altering decisions that affect teachers, administrators, and students are made on the basis of testing. Tests determine who is promoted and who is retained. Tests can determine who will receive a high school degree and who will not. Test scores can determine if a school will be reconstituted, if there will be job losses for teachers and administrators when scores do not improve, or if there will be cash bonuses when scores do improve. Thus, there is no question as to the emphasis and importance placed on a single test score. Under these conditions, we must worry that the process that is being monitored by these test scores—quality of education—is corrupted and distorted simultaneously, rendering the indicator itself less valid, perhaps meaningless.

CONCLUSION

In George W. Bush's state of the union address in January 2006, he proclaimed that America was addicted to oil.⁴² He might have noted that America is also addicted to television, fast food, and testing. But these are not real addictions; they are socially created ones, and as such they can be changed through policies that are more sensible than those in place now. Slowly, over the course of the past two or three decades, Americans have been fed the increasingly urgent propaganda that American education is about to collapse. We have been led to believe that if something didn't change soon, we would have a crisis on our hands. It is no wonder our citizenry wants action. We have been led to believe in the "standardized tests as god" religion by the business and test-manufacturing community.⁴³ In the absence of any visible or viable alternatives, high-stakes testing is presented as the only possible answer to thwart our descent into mediocrity. But opposition to these messages exists.

In the years since NCLB was passed, the number of critics of high-stakes testing and other aspects of the legislation has grown substantially.⁴⁴ This growing outcry has prompted federal policymakers to make some changes to NCLB—specifically with respect to special-education and second-language-learner populations. But it isn't enough. In the pages that follow, we argue that the continued use of high-stakes testing as a mechanism for driving school change is fundamentally flawed and that serious damage is being inflicted on schools and our children.

The government did not heed the warnings of 9/11. Neither did they heed the warning of Katrina, the hurricane that wreaked havoc on the people of New Orleans. Now they do not heed the warnings about the social security and medical trust funds that indicate things will surely reach crisis proportions in a few more years. So we are not sanguine about our own warning that the schools are suffering as a function of high-stakes testing. But we do know that we live in a democracy and that concerted action by our citizenry brings about change. We hope this book will spark such action.