



Figure 1.1. Timeline of Major Dates

CHAPTER 1

A BRIEF HISTORY OF ACCOUNTABILITY AND STANDARDIZED TESTING

In “Harvard, Again: Considering Articulation and Accreditation in Rhetoric and Composition’s History,” Ryan Skinnell demonstrates how the demand for articulation and its counterpart, accreditation, as larger institutional processes, have had profound effects on the development of writing instruction in the United States. He defines articulation as “the institutional protocol for connecting two or more types of schools (for example, secondary and post-secondary) so that students can move between them by virtue of well-defined processes” (96). In particular, he traces the impact of Harvard’s need for improved articulation from high school to Harvard, and the resulting accreditation practices, as a significant shaping force on rhetoric and composition in the contemporary academy (49). Skinnell’s work, in essence, positions these accreditation practices as an early example of the accountability measures increasingly at play in today’s K–12 classrooms.

While Skinnell focuses narrowly on articulation and accreditation in relation to one particular institutional context, our investigation extends this view, looking at how calls for *accountability* contribute to the current rhetoric of crisis framing our national discourse on writing instruction. More specifically, we track how calls for accountability have increasingly resulted in the use of standardized tests of writing, despite little evidence that increased use of standardized testing over the last 150 years has led to improved writing or improved measurement of college readiness.

In fact, the most recent National Assessment of Educational Progress writing assessment reveals that only 24 percent of our eighth and twelfth graders performed at the proficient level, with just over half of them determined to perform at the basic level (partial mastery, not college-ready writers) (National Center for Educational Statistics 2012). This is despite many years of teachers laboring under the standardized testing requirements mandated by the No Child Left Behind Act. Further, George Hillocks Jr.’s extensive study of writing assessment in five states—Illinois, Kentucky, New York, Oregon, and Texas—“indicates that many writing assessments do not have the intended effects . . . [and] what they [teachers] are teaching appears to have a negative effect on the students in states with poorly thought out assessments” (205). In terms of standardized tests as a predictor of college readiness, we can point to a recent study released by the

National Association for College Admission Counseling—“Defining Promise: Optional Standardized Testing Policies in American College and University Admissions.” This study asks whether or not standardized testing produces valuable predictive results, or if it limits the pool of applicants “who would succeed if they could be encouraged to apply” (Hiss and Frank 2). Examining data for 123,000 students at twenty private colleges and universities, six public universities, five minority-serving schools, and two arts schools, the researchers found that “the differences between submitters [of ACT or SAT scores] and non-submitters are five one-hundredths of a GPA point, and six-tenths of one percent in graduation rates. By any standard, these are trivial differences” (3). Equally important is the finding that those students who don’t submit test scores are more likely to be first-generation, minorities, women, Pell Grant recipients, and students with learning differences (3).

Understanding the path that led to accountability through standardized testing is especially important for those of us working in higher education at this moment in time, as the use of standardized tests for students already in college (e.g., rising junior) increases every year.¹ Complicating matters further, two tests originally designed to measure student mastery of Common Core State Standards (CCSS),² PARCC (Partnership for Assessment of Readiness for College and Careers) and SBAC (Smarter Balanced Assessment Consortium), are also being adopted by some states for college placement and admissions.³ For example, the Illinois State Board of Education’s “PARCC Assessment FAQs” page states that “Institutions of higher education are working toward acceptance of PARCC assessment results as a way to show readiness for college-level work without remediation . . . allowing colleges and universities to place those students testing at the ‘college and career ready’ level in credit-bearing courses (as opposed to remedial courses).” The Colorado Department of Education’s Communication Division Assessment Fact Sheet states online that “Colorado’s new [2014] higher education admissions and remediation policies allow institutions to use PARCC scores for both course placement and admissions purposes.”

As will become evident in this chapter, the recurring calls for educational reform that shape so much of our national discourse include standardized testing as one of the primary ways of enacting reform. These calls have more often than not resulted in increased use of standardized tests despite almost one hundred years of published debate and little consensus as to how well standardized testing can measure and improve educational performance. It is important to recognize that even though the results of standardized tests tell us very little about actual classroom practices, they have become the most visible and widely available measure of our classrooms. Finally, our exploration of calls for accountability through standardized testing strives to lay bare increasingly well-funded systems

of control over our classrooms, as well as to deepen our understanding of emerging forces so that we can chart a path forward.

STANDARDIZED TESTING BEFORE SPUTNIK

While we often associate standardized testing of writing achievement with contemporary K–12 classrooms, standardized testing has long been commonplace in the United States. Most popular accounts identify the late 1960s as the beginning of *the* age of standardized testing, coming on the heels of America’s perceived loss of the space race to Russia with the launch of Sputnik. For example, in 1969 Alice Rivlin, who is still considered one of America’s leading economists, was asked to participate in a conference titled “The Measurement of Economic and Social Performance” (the proceedings of which were later published by the National Bureau of Economic Research). The planning of this conference coincided with the birth of the National Assessment of Educational Progress (NAEP), the first U.S. national test of academic achievement. Rivlin’s task was to address the measurement of performance in education. She wrote:

This is the age of testing. Considerable effort has gone into devising and standardizing a wide variety of tests of intellectual skills and accumulated knowledge. Billions of man-hours of student and teacher time are devoted to taking, administering, grading, analyzing, and discussing standardized tests. One might hope that all the effort would tell us something about output or performance in education.

Remarkably, almost no information presently exists which would give a basis for constructing an index of change in educational test scores over time. (423)

We generally agree with Rivlin’s characterization of the state of educational measurement at that time, but it is important to point out that standardized testing has a much longer and varied history than her paper and many accounts of the educational reform movement would make it seem. In fact, the late 1960s/early 1970s stand as just one of many periods during which standardized testing was a central measure of cultural and economic capital throughout history.

The 1860s mark the beginning of a visibly significant change in the history of U.S. universities as large numbers of students from disparate backgrounds began to seek a university education, defined academic disciplines became the organizing principle, and a professorate emerged that more closely resembles that of today. During this same time, Congress established the Department of Education (1867). Succumbing to intense pressure to keep the federal government

out of what many saw as the province of the states, Congress demoted it to an Office of Education in 1868. The Office of Education spent time being shuffled between the Department of the Interior and the Federal Security Agency before settling in the Department of Health, Education, and Welfare. It was eventually given cabinet-level authority as the Department of Education in 1980 (U.S. Department of Education, *The Federal Role in Education*).

The main purpose of establishing the Department of Education in 1867 as described in the Act was to have an agency that gathered information on the condition and progress of our educational system:

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, That there shall be established, at the city of Washington, a department of education, for the purpose of collecting such statistics and facts as shall show the condition and progress of education in the several States and Territories, and of diffusing such information respecting the organization and management of schools and school systems, and methods of teaching, as shall aid the people of the United States in the establishment and maintenance of efficient school systems, and otherwise promote the cause of education throughout the country. (An Act to establish a Department of Education, 1867)

Upon establishment of this department, a number of people began to advocate for the implementation of a national standardized exam. But a strong adherence to states' rights and logistical barriers to the implementation of a national exam kept such efforts at bay. In fact, it would be just over one hundred years after establishing a Federal Department of Education, in 1969, that the first national exam, the NAEP, was administered. Nonetheless, the mid-1800s and early 1900s marked a rapid expansion and development of educational testing and measurement in the United States—much of it through the efforts of our universities to influence curriculum at the secondary level in order to ensure students were prepared for university-level work. While these early efforts were not referred to in terms of accountability, but rather as a process of articulation, this process helped lay the groundwork for the systems of accountability at play today.

As early as 1833 Harvard and other colleges began to administer written exams as proof of achievement—the first in math (Black 192). By 1851, Harvard faculty recognized they could no longer assume students would arrive with a uniform set of skills, and in response instituted one of the first standardized, written entrance exams, focusing primarily on Latin grammar and math (Hanson 193) and, by the mid-1860s, including Greek composition, history, and ge-

ography. During this same time period, the number of children in government-funded schools began to swell, and public schools began to follow the example set by colleges in terms of measuring achievement. With increasing demand from universities for these schools to produce college-ready students, as well as the organization of boards of education in the states, standardized testing began to find solid footing in the United States.

The written standardized exam administered to all Boston school children in 1845 is thought to be the first large-scale achievement test of its kind, and a full account of this exam, including test questions, sample responses, and results, was collected for the 1925 edition of *Then and Now in Education, 1845:1923* (Caldwell and Courtis). Prior to 1845, the Boston public schools followed the standard practice of requiring oral exams administered by a traveling panel of examiners. But by 1845 there were 7,000 students in nineteen different schools, and this approach to measurement was no longer feasible. Instead, Boston instituted a written exam thought to be more objective, reliable, and economical than the oral exams (Mathison 3). The language sections of these tests focused on definitions and prescriptive grammar. Early examiners describe the condition and progress of education in Boston schools at this time in their report:

The first feeling occasioned by looking over these returns is that of entire incredulity. It is very difficult to believe that, in the Boston Schools, there should be so many children in the first classes, unable to answer such questions; that there should be so many who try to answer, and answer imperfectly; that there should be so many absurd answers, so many errors in spelling, in grammar, and in punctuation. If by any accident these documents should be destroyed, we could hardly hope that your faith in our accuracy would induce you to believe the truth if we told it. But the papers are all before you, each signed by the scholar who wrote it. . . . The most striking results are shown in the attempts to give definitions to words. There were twenty-eight words selected from the reading book, which the classes have probably read through during the year, and some probably more than once. Some of these words are the very titles or headings of reading lessons; some of them occur several times in the book, and yet, of the 516 children who had these questions before them, one hour, not a single one defined correctly every word; only 47 defined half of them; and 29 could not define correctly a single one of the whole 28 words. (*Then and Now* 171, 175)

While most of these very early tests did not resemble those with which we are familiar today, it was not long before the basic structure of standardized tests of written communication were in place—a structure to which we still largely adhere. Standardization of writing tests took a significant leap forward in 1860 with the introduction of scaled tests of writing achievement. George Fisher, an English schoolteacher, provided us with the first written account of educators using anchor papers on a scale of 1–5 designed to measure writing achievement of large numbers of students. Fisher used these tests to assess handwriting, spelling, grammar, and composition (Bryant and Bryant 420). While it is not clear if the standard scale books themselves still exist, Fisher’s description of them can be found in a copy of a paper he presented to the Statistical Section F, British Association, Cambridge, October 1, 1862:

**On the Numerical Mode of Estimating and Recording
Educational Qualifications As Pursued in the Greenwich
Hospital Schools**

It has been observed that “no mode of teaching can be properly appreciated so long as we are without recognized principles of examination, and accuracy in recording the results; for without such means neither failures nor improvements will add to our common stock of experience in such matters; and we hand down to posterity no statistical information of such value as will mark the progress of Education. . . .

Such a plan of numerical estimation has been carried out in the Greenwich Hospital Schools. A book, called the “Standard Scale-Book,” has been there kept since the first general introduction of the plan containing the numerical value of each degree of proficiency in the various subjects of examination. If it be required, for instance, to determine the numerical equivalent to any specimen of writing, a comparison is made with various standard specimens of writing contained in this book, which are arrayed and numerically valued according to the degree of merit. The best executed being represented by the number 1, and the worst by the number 5. . . . So long as such standard specimens are preserved in the School, constant and permanent values for proficiency in writing can be maintained; and since facsimiles can now be multiplied with very little expense, it appears obvious that the same principle might be generally adopted, provided well-considered standards were agreed upon and recognized. . . .

I trust that I have made this mode of Registration as intelligible to the Section as the present circumstances will admit of. I have no other motive in making this communication to them, beyond the desire of exciting the attention of others to the subject, that it may lead to the adoption of some sound practical system of testing and recording educational qualifications . . . according to some fixed standards of valuation as might be generally agreed upon by those engaged in Education. (Excerpted from Cadenhead and Robinson 16–18)

Following Fisher's efforts to develop a standardized instrument to measure writing achievement is the 1912 development of the Hillegas-Thorndike Scale for the Measurement of Quality in English Composition by Young People. Dr. Hillegas, a professor at Columbia University, believed that uniform standards would establish a level of certainty when comparing the work of a student, school, and system of schools with that of others. Further, if these measurements of school performance: "approximate[ed] the accuracy of the scales used in measuring extension, weight, and time, educational administrators and investigators would be able to measure and express the efficiency of a school system in terms that would carry conviction" (2).

Development of the Hillegas-Thorndike Scale involved hundreds of people, with final judgments resting in the hands of "28 individuals, nine of whom were 'men of special literary ability,' eleven 'gifted teachers familiar with secondary education,' and eight 'psychologists familiar with the significance of scales and zero points in the case of intellectual abilities and products'" (Johnson 40). Beginning with a sample of 7,000 essays divided into ten levels of ability, these educational researchers developed a scale from 0–93 that eventually included twenty-nine samples or anchor papers that were designed to allow for measurement of the "absolute gain which any pupil made in any year . . . the same as his gain in height, weight, wages or pulse rate and the results of different means and methods of teaching could be demonstrated with exactitude instead of being guessed at" (Thorndike 214). We might consider this a very early type of the "value-added assessments" that form the basis of many a "new" reform, including pay-for-performance teacher contracts and the use of growth models. Also interesting is that—whether rooted in a rubric or the use of model anchor papers (or both), whether based on evaluation by local teachers, experts, or a software program, and whether on a small scale using performance-based portfolios or a large scale using a spiraling, balanced incomplete design (e.g., NAEP)—scaled writing assessment is still the most common type of writing assessment used at both local and national levels. In other words, when it comes

to assessing writing, we are still using a system originally developed in the 1860s in England and then later refined in the early 1900s in the United States.

The Hillegas-Thorndike Scale, and the goals of Hillegas and Thorndike themselves, were widely debated in composition teaching and research publications of the time, including numerous references from 1912–1925 in NCTE’s *English Journal*. While many found the scale useful in very controlled contexts, most found it impractical due to the variation among genres, styles, grade levels, and other matters familiar to us today. As one critic pointed out, “You can not measure light, and warmth, and redness on the same rod” (Thomas 3) and, similarly, you can not measure all student writing achievement using the same rod (Thomas 3). Even in the twenty-first century, with technology unimagined in the early twentieth century, we are still using the same rod to measure student writing achievement. Rather than use technology to bring wide-scale innovation to this process, we have been content to focus on bringing economy of scale to the process.

One other major development requires mention in our brief history. In 1900, the College Entrance and Examination Board (CEEB) was established by a group of private high schools and elite colleges in order to standardize the admissions process and drive a more uniform curriculum at the private New England high schools from which the colleges drew most of their students. The CEEB later became College Board, a nonprofit testing agency most of us are familiar with as the administrator of the SAT. By the mid-1950s College Board was administering the Advanced Placement Program and soon developed the PSAT to measure students’ critical reading and math skills in preparation for college entrance exams like the SAT and ACT. In 1959 ACT was formed as an alternative testing option to the SAT. Both of these organizations have grown immensely over the years, reaching ever farther into the educational landscape.

SPUTNIK: A MOMENT OF CRISIS

We’ve now returned to the point in our history where popular accounts of standardized testing and educational reform generally begin—Sputnik. Sputnik was the first artificial Earth satellite launched by the Soviet Union in 1957, signaling America’s advertised loss of the space race. Homer Newell, a theoretical physicist and mathematician at the Naval Research Laboratory as well as NASA historian, recalls the moment:

How brightly the Red Star shone before all the world in
October of 1957! Streaking across the skies, steadily beeping
its mysterious radio message to those on the ground, Sputnik
was a source of amazement and wonder to people around

the globe, most of whom had had no inkling of what was about to happen. To one nation in particular the Russian star loomed as a threat and a challenge.

In the United States many were taken aback by the intensity of the reaction. Hysteria was the term used by some writers, although that was doubtless too strong a word. Concern and apprehension were better descriptions. Especially in the matter of possible military applications there was concern, and many judged it unthinkable that the United States should allow any other power to get into a position to deny America the benefits and protection that a space capability might afford. A strong and quick response was deemed essential. (Mudgway 75)

Many have and continue to use this event as proof of declining educational standards, particularly in math and science, making ample room for the argument that education is a matter of national security and the common good, and thus requires federal intervention. But Sputnik may be the most successful and persistent manufactured myth about the state of America's educational system to date. The crisis generated by this manufactured myth allowed for the political capital needed to pass the National Defense of Education Act⁴ in 1958, opening the door to a national test of achievement—the NAEP—a giant leap toward the accountability movement that is now in full swing. Furthermore, this is a crisis that has remained a persuasive touchstone for educational reform movements for almost sixty years.

For example, Christopher Tienken and Donald Orlich remind us:

President Bill Clinton's Secretary of Education Richard Riley (1995) used Sputnik to justify further federal involvement in education as part of the America 2000 legislation: "When the Russians woke us up by flying Sputnik over our heads late at night—a few of you may remember that experience—Congress passed the 1958 National Defense of Education Act, which sent millions of Americans to college and educated a generation of scientists who helped us to win the Cold War." Ronald Reagan used Sputnik as a propaganda tool in 1982 to support his plan to give tax credits for parents to send their students to private schools. (25)

And in his 2011 State of the Union Address, President Obama declared: "This is our generation's Sputnik moment. . . . But if we want to win the fu-

ture—if we want innovation to produce jobs in America and not overseas—then we also have to win the race to educate our kids. That’s why instead of just pouring money into a system that’s not working, we launched a competition called Race to the Top.”

But Tienken and Orlich assert that recently declassified documents from the Eisenhower administration “tell another story of Sputnik. Sputnik became a manufactured crisis, to borrow a term by Berliner and Biddle” (21). It’s important to keep in mind that, at this time, the federal government had very little to do with our K–12 curriculum, and attempts by the federal government to shape curriculum were easily rebuked. Sputnik was quickly framed as tangible, startling evidence of a broken educational system, and a crisis of opportunity ensued. But there is significant evidence showing that the launch of Sputnik had nothing to do with the state of our educational system. For example, a memorandum of conference with President Eisenhower on October 8, 1957, quotes then Deputy Secretary of Defense Quarles, as saying, “the Redstone [military rocket] had it been used could have orbited a satellite a year or more ago” (qtd. in Tienken and Orlich 21).

NASA’s own history of Charles Pickering, director of the Jet Propulsion Laboratory from 1954 to 1976, also tells a story of Sputnik much different than that continually forwarded by our leaders when arguing for school reform efforts. Pickering’s own account of the nation’s reaction to Sputnik reveals the sense of helplessness and urgency this event elicited:

The reaction in this country was amazing. People were startled to realize that this darn thing was going overhead about ten times per day and there was not a thing they could do about it—and realizing that what was thought to be a nation of peasants could do something like this—with this amount of technical complexity. (qtd. in Mudgway 75).

Pickering and his staff at the Jet Propulsion Laboratory not only knew how to launch a satellite into space, they had all the hardware they needed in order to do it. “All they lacked was the approval to ‘go ahead.’ . . . But the word to ‘go ahead’ did not come” (Mudgway 75). It is beyond the scope of this book to explore the reasons why the United States held off on its launch of a satellite—many of which center on concerns about the state of the Cold War at the time. Of importance to our discussion, one of the unexpected results of the decision to delay the launch was the passage of the National Defense of Education Act, the first federal policy to largely target higher education while also directing funds to improving instruction in math, science, and modern foreign languages (e.g., Russian) in our K–12 classrooms.

THE PRODUCT OF AN EXTENSIVE SURVEY REQUESTED BY THE CIVIL RIGHTS ACT OF 1964, THIS REPORT DOCUMENTS THE AVAILABILITY OF EQUAL EDUCATIONAL OPPORTUNITIES IN THE PUBLIC SCHOOLS FOR MINORITY GROUP NEGROES, PUERTO RICANS, MEXICAN-AMERICANS, ORIENTAL-AMERICANS, AND AMERICAN INDIANS, AS COMPARED WITH OPPORTUNITIES FOR MAJORITY GROUP WHITES. COMPARATIVE ESTIMATES ARE MADE ON A REGIONAL AS WELL AS ON A NATIONAL BASIS. SPECIFICALLY, THE REPORT DETAILS THE DEGREE OF SEGREGATION OF MINORITY GROUP PUPILS AND TEACHERS IN THE SCHOOLS AND THE RELATIONSHIP BETWEEN STUDENTS' ACHIEVEMENT, AS MEASURED BY ACHIEVEMENT TESTS, AND THE KINDS OF SCHOOLS THEY ATTEND. EDUCATIONAL QUALITY IS ASSESSED IN TERMS OF CURRICULUMS OFFERED, SCHOOL FACILITIES SUCH AS TEXTBOOKS, LABORATORIES, AND LIBRARIES, SUCH ACADEMIC PRACTICES AS TESTING FOR APTITUDE AND ACHIEVEMENT, AND THE PERSONAL, SOCIAL, AND ACADEMIC CHARACTERISTICS OF THE TEACHERS AND THE STUDENT BODIES IN THE SCHOOLS. ALSO IN THE REPORT IS A DISCUSSION OF FUTURE TEACHERS OF MINORITY GROUP CHILDREN, CASE STUDIES OF SCHOOL INTEGRATION, AND SECTIONS ON HIGHER EDUCATION OF MINORITIES AND SCHOOL NONENROLLMENT RATES. INFORMATION RELEVANT TO THE SURVEY'S RESEARCH PROCEDURES IS APPENDED. NOTABLE AMONG THE FINDINGS ON THE SURVEY ARE THAT NEGRO STUDENTS AND TEACHERS ARE LARGELY AND UNEQUALLY SEGREGATED FROM THEIR WHITE COUNTERPARTS, AND THAT THE AVERAGE MINORITY PUPIL ACHIEVES LESS AND IS MORE AFFECTED BY THE QUALITY OF HIS SCHOOL THAN THE AVERAGE WHITE PUPIL. THIS DOCUMENT IS ALSO AVAILABLE FROM THE SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, D.C. 20402, FOR \$4.25. (AH)

Figure 1.1 Abstract from Coleman Report

But it wasn't just the perceived loss of the space race that finally led to a national exam. *Equality of Educational Opportunity for All*, often referred to as the Coleman Report, may have had an equally important effect. The Coleman Report was commissioned by the U.S. Office of Education and published in 1966. It "marks the first time there is made available a comprehensive collection of data gathered on consistent specifications throughout the whole nation" (Coleman 1). Approximately 645,000 students from 4,000 public schools in grades 3, 6, 9, and 12 participated in this research, which focused on the extent to which equality of education was a reality for America's school children.

This was a landmark study leading to a flurry of activity but, as many argue, little in the way of educational progress. As a brief aside, we can link The Colman Report with claims such as those of Berliner that the real education crisis is a crisis of poverty, not a crisis of overall educational achievement. In a retrospective on The Coleman Report, Adam Gorman and Daniel Long of the Wisconsin Center for Education Research found that forty years later the major findings of the report hold up well, most notably that per-pupil spending is less important than level of teacher training, the black-white achievement gap persists, and "Student achievement still varies substantially within schools . . . and

this variation is still tied to students' social and economic backgrounds" (19). In fact, when discussing the 2015 reauthorization of the Elementary and Secondary Education Act, Secretary of Education Arne Duncan prioritized equity for low-income and minority students because "Education Department data show that 6.6 million students from low-income families are being shortchanged when it comes to state and local education funding" (U.S. Department of Education, "Secretary Duncan"). For example, the education department estimates that in Pennsylvania, the highest-poverty districts spend 33 percent less than the lowest-poverty districts, while in Vermont, Illinois, Missouri, and Virginia, the highest-poverty districts spend 18–17 percent less than the lowest-poverty districts. And in Nevada, the highest-minority districts spend 30 percent less than the lowest-minority districts, while in Nebraska and Arizona, the highest-minority districts spend 17–15 percent less than the lowest-minority districts (U.S. Department of Education, "Secretary Duncan").

Importantly, it is difficult to attract, retain, and develop high-quality teachers in high-poverty schools (Clotfelter, Ladd, and Vigdor 2005; Grissom 2011). A 2014 report by the Alliance for Excellent Education estimates that 13 percent of our teachers move or leave the teaching profession each year: "This high turnover rate disproportionately affects high-poverty schools and seriously compromises the nation's capacity to ensure that all students have access to skilled teaching" (Haynes). This is especially problematic when we consider that, as Ben Ost says, "one of the most consistent findings in the literature on teacher quality is that teachers improve with experience" (1).

Most studies of teacher turnover in high-poverty schools have attributed turnover to characteristics of the students and the teachers, rather than the organizational structure of the schools themselves—organizational structures that can be improved with increased funding. Emerging research on teacher turnover in high-poverty schools suggests "when these teachers leave, it is frequently because the working conditions in their schools impede their chance to teach and their students' chance to learn" (Simon and Johnson 4). Organizational factors that are associated with higher rates of turnover include administrative support, teacher input in decision-making, salary, and aspects of school culture (Simon and Johnson 12). We will return to a discussion of some of these factors in our last chapter, but for the moment let's turn back to our history of standardized testing.

The ability tests collected as part of The Coleman Report were administered by ETS and the language section focused on items such as sentence completion and identifying analogies—items that could easily and efficiently be measured. This is not surprising given the number of students involved in this study and research appearing as early as the 1940s claiming a high correlation between

objective tests (tests of grammar, spelling, punctuation, and capitalization) and final grades in rhetoric courses. For example, Irwin A. Berg, Graham Johnson, and Robert P. Larsen detail a study conducted in 1943 wherein the researchers agree that under ideal conditions writing proficiency can only be determined by a demonstration of writing, but also argue that an objective test is correlated highly enough with final grades in rhetoric courses that “the advantages of rapid scoring which could be done by persons who are not necessarily rhetoric instructors, together with the advantages of objectivity of score, would make the use of a suitable objective test an extremely practical measuring tool” (430). Of further note, the objective test used in this study was scored by an International Business Machines electrical scoring machine, a machine first introduced in 1937 that allowed for scoring of large numbers of standardized tests on a new scale. Much like the history of standardized tests of writing instruction, arguments for the use of machine-graded scoring to assess writing ability have a much longer history than many of our current discussions reveal. And, it is these histories that we must begin to more fully understand if we are to reassert the agency of teachers and students in the current clash over the control of literacy.

While The Coleman Report was intended as a massive, one-time educational measurement and analysis effort, the NAEP writing assessment, also known as *The Nation's Report Card*, was intended to be repeated on a regular basis, allowing for comparison of educational progress over time. The NAEP tests students in different subjects each year, with a writing test first administered in 1969/70 and repeated approximately every four years. The overarching goal of NAEP, as stated in the 1969/70 writing report, is to assess educational attainment on a national basis; it is also to offer “descriptions of what groups of Americans know and can do and, over a period of time, of whether there is progress in educational attainments” (1). Finally, Americans would know how our students are performing not just during any given year, but also over time, so that we could track educational progress. While this plan has worked relatively well for reading and math, the same two subjects mandated for yearly assessment by NCLB, it has not worked for writing. Instead, teachers and administrators who have looked at the results of such tests for guidance are often left confounded. An account written in 1992 by Mary Licklider, a junior high English teacher, is representative:

The Nation's Report Card on Writing issued by the National Assessment of Educational Progress (NAEP) in June 1990 left me frustrated and confused. I could not tell whether students' writing had declined or improved since 1970. From the tone of the report I suspected the former. As an English teacher, I thought I might be more effective selling shoes . . . Surely, the

extensive resources of NAEP, including a massive data bank covering two decades, might yield information that teachers need if they are to become better teachers of writing. I feel somewhat short changed by the reports I have read; and I have been unable to obtain essential NAEP documents even with the help of interlibrary loan operating through our local public library and reaching out of state as well. (34, 39)

In fact, NAEP did attempt to produce long-term trend reports for writing, but by 1999 had abandoned this effort, explaining that the content and manner of administration had changed so much from test to test that the accuracy of the results were called into question (Phillips). Curiously, while it is possible to track down these reports via used bookstores and microfiche, they can no longer be easily accessed online via the Education Resources Information Center (ERIC) or the NAEP website.

The problem of large-scale writing assessment is stated very clearly by Arthur Applebee, drawing on a paper he was commissioned to write by the National Assessment Governing Board (overseers of the NAEP since 1988) in 2005 as they worked to develop a framework for the 2011 NAEP writing assessment:

Underlying all of the specific issues that follow is a larger one: What information about how students write should NAEP and other large-scale assessments provide to interested members of the general public, policymakers, and educators? Although it is a seemingly simple question, buried within it are a variety of difficult issues on which there is currently little consensus, including how to describe the domain of writing tasks; the relationships among component skills, content knowledge, and generalized writing “fluency”; and the relevance of computer-based applications to definitions of writing achievement as well as to assessment techniques. (“Issues” 82)

In other words, writing is an extremely complex and ever-changing human activity, continually influenced by evolving cultural norms and technological advancements. Pinning it down for large-scale assessment over time simply hasn't been possible. If an examination of long-term trends using the only large-scale, longitudinal studies publicly available teaches us anything, it is how exceedingly difficult it is to measure the writing achievement of students in rigorous and valid ways over any extended period of time using a single measurement tool such as a standardized test—especially in ways that can guide instruction. This is not to say that a national assessment of writing instruction isn't useful, but rather

that a test has yet to be developed that can reliably measure change in writing achievement over time due to the rapidly changing writing demands placed on students and workers.

Despite the misgivings of the National Assessment Governing Board itself, and constant revision and critique of attempts to assess writing over time, many were not deterred by these concerns and instead began to argue for the use of such tests within higher education.

STANDARDIZED TESTING AND HIGHER EDUCATION

We began this chapter with a section from Alice Rivlin's presentation at the Measurement of Social and Economic Performance Conference in 1969. After arguing that most aspects of educational performance can and should be measured, Rivlin concludes:

Test scores and other performance measures are now being used as evidence against educators. It seems likely that educators will respond by developing more comprehensive and reliable measures of their own, not only to satisfy the public, but to put their own house in order and build into the management of education some measures of what is being produced and some incentives to produce it more effectively. (427–28)

Within higher education, it is the case that many educators and the professional organizations that represent them responded by developing comprehensive and reliable measures of their own. For example, the National Institute for Learning Outcomes Assessment, established in 2008 and located in the School of Education at the University of Illinois at Urbana-Champaign, holds as its mission to “discover and disseminate ways that academic programs and institutions can productively use assessment data internally to inform and strengthen undergraduate education, and externally to communicate with policy makers, families and other stakeholders.”⁵ The multidimensional toolkit they propose includes tests, surveys, portfolios, curriculum mapping, benchmarking, handbooks, and rubrics.

As another example, the Peer Review of Teaching Project (PRTP), begun in 1994 and currently housed at the University of Nebraska-Lincoln

is a faculty-driven approach for developing a campus climate for teaching improvement and reform. Invited faculty work in teams over the course of a year to discuss approaches for documenting and assessing student learning within particular

courses. Rather than advocating any particular teaching approach or technique, the PRTP focuses on helping faculty document student learning occurring in their course and then reflect on whether student performance demonstrates achievement of the curricular and department goals.⁶

Specifically in relation to writing classrooms, we might look to the National Council of Teachers of English and the Council of Writing Program Administrators' (NCTE-WPA) "White Paper on Writing Assessment in Colleges and Universities, the Conference on College Composition and Communication's (CCCC) "Writing Assessment Principles," the collaboration between WPA and the National Survey of Student Engagement, and the Valid Assessment of Learning in Undergraduate Education (VALUE) Rubric for Written Communication offered by the Association of American Colleges and Universities through their VALUE Rubric Development Project. All of these efforts propose pedagogically sound, empirically based assessment practices. However, for multiple reasons, these efforts have not satisfied the public or deterred calls for more standardized testing and accountability. Instead, we have found ourselves in a defensive position, as evidenced by the establishment of the NCTE Rapid Response Assessment Task Force in 2014. Led by Kathleen Yancey, this task force was created "to address the growing cacophony around assessment" from a very activist stance.

In the remainder of this chapter, we will look at two defining texts and one potentially major shift in public policy agenda setting in the call for more standardized testing at the college level that epitomize the need for work such as that of the NCTE Rapid Response Assessment Task Force. The two texts are *A Test of Leadership: Charting the Future of U.S. Higher Education* (a report of the commission appointed by Secretary of Education Margaret Spellings, also known as the Spellings Report) and one of the most widely discussed books calling for reform of higher education, *Academically Adrift* (Arum and Roksa). Examining these texts within the context of their history increases our understanding of present and emerging forces so that we can chart a path forward. We will conclude with an exploration of the shift toward advocacy philanthropy and the emerging role of foundations in directing policy and practices in U.S. higher education.

As explained in our introduction, our current work is motivated by many happenings in 2006, including efforts to set higher education on the same path as K–12 through the No Child Left Behind Act. It is the 2006 report commissioned by then Secretary of Education Margaret Spellings that aimed to establish systems of accountability, punishment, and reward not seen before in higher education. Echoing the same threats to the American Dream and the Common

Good as its predecessors, the National Defense of Education Act and *A Nation at Risk*, the U.S. Department of Education's *A Test of Leadership: Charting the Future of U.S. Higher Education* urges a "robust culture of accountability" (20):

We believe that improved accountability is vital to ensuring the success of all the other reforms we propose. Colleges and universities must become more transparent about cost, price, and student success outcomes, and must willingly share this information with students and families. Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a "value-added" basis that takes into account students' academic baseline when assessing their results. This information should be made available to students, and reported publicly in aggregate form to provide consumers and policymakers an accessible, understandable way to measure the relative effectiveness of different colleges and universities. (4)

Interestingly, and very much in line with the rhetoric and practice of No Child Left Behind, the authors of this report note in their introductory summary that, "According to the most recent National Assessment of Adult Literacy . . . the percentage of college graduates deemed proficient in prose literacy has actually declined from 40 to 31 percent in the past decade" (3). And yet, in its recommendations, the commission "urge[s] these institutions to develop new pedagogies, curricula and technologies to improve learning, *particularly in the areas of science and mathematics*" (5, emphasis ours), choosing not to place an emphasis on writing in U.S. schools.

A Test of Leadership names specific standardized tests, such as the Collegiate Learning Assessment (CLA), for use in our colleges as a means of rigorous accountability. The CLA was developed under the auspices of the Council for Aid to Education (CAE), a nonprofit organization initially established in 1952 to encourage corporate support of education. The CAE currently conducts policy research on higher education as well as focuses on improving quality and access in higher education, primarily through the CLA, and now CLA+ (a revision of CLA). CAE describes CLA+ as a way for national and international institutions to "benchmark value-added growth in student learning at their college or institution compared to other institutions." CAE uses "performance-based tasks . . . to evaluate the critical-thinking and written-communication skills of college students. It measures analysis and problem solving, scientific and quantitative reasoning, critical reading and evaluation, and critiquing argument, in addition

to writing mechanics and effectiveness” (“CLA+ Overview”). Our primary concern here is not with the CLA itself. Although we do find some of the claims about the use and value of CLA to be problematic, as we explain later, within clearly defined and well-understood parameters it can be a useful tool, although only when combined with other measures of educational progress.

Instead, our primary concern with the CLA is the way that it is employed in the name of accountability, following the example of one of the most popular books on higher education today—*Academically Adrift*—which has been touted by those on the right, left, and center as proof of a failing system of higher education in need of unprecedented levels of control and accountability. As noted by many researchers, it took just over a year for the central touchstone of this book, that 45 percent of the students in the study failed to show significant gains in reasoning and writing skills between the beginning of their freshman year and the end of their sophomore year, to establish itself as central tenet of U.S. folklore about higher education (see, e.g., Astin, Lane and Oswald). Like many others, we have great concerns about the statistical analysis in *Academically Adrift* and its sweeping claims based on a study of only slightly more than 2,300 college students. But perhaps more importantly, from our perspective as writing researchers, we have serious concerns about the claims made by CLA that it is testing “general” reasoning and writing skills and, furthermore, that the results of a 90-minute performance-based task can measure the writing abilities of students over time.

The authors of *Academically Adrift*, with the help of CLA, enlisted twenty-four colleges of varying sizes and classifications to participate in their research. It is important to note that they do not include community college students in their research sample, and further, of the 2,300 students who volunteered to participate in this study, very few were considered to be of low scholastic ability. This is important because community college students and students of lower scholastic ability are likely to exhibit the most growth during their college career. Further, this book focuses on the results of just one performance task, giving students “ninety minutes to respond to a writing prompt that is associated with a set of background documents” (21).

The details of the statistical errors made by the authors have been explained by many researchers. For example, the authors set the level of statistical significance at .05—a relatively arbitrary starting point. Using this number, they claim that 45 percent of the students in this study did not improve their reasoning and writing skills because the overall change in scores was not statistically significant. As Alexander Astin points out in the *Chronicle of Higher Education*, “Just because the amount of improvement in a student’s CLA score is not large enough to be declared ‘statistically significant’ does not prove that the student failed to

improve his or her reasoning and writing skills” (4). In fact, as Richard Haswell makes clear, “every one of their twenty-seven subgroups recorded gain” (488), but the authors of *Academically Adrift* claim that this gain was “modest” or “limited” based on their set standard of statistical significance. Equally concerning, as Haswell explains, “Not one piece of past research showing undergraduate improvement in writing and critical thinking—and there are hundreds—appears in the authors’ discussion or their bibliography, although both are a swim with think-tank books and blue-ribbon papers opining the opposite” (488).

Examined from another angle, Lane and Oswald make the case that:

This 45% finding is, indeed, shocking—but for a completely different reason. Considering that each significance test was based on a sample size of 1 (i.e., each student’s change in the CLA measure), it is hard to imagine that as many as 55 percent of students would show statistically significant gains. Indeed, one would expect to find an order of magnitude fewer significant improvements, based on the mean difference between the pre- and post-tests the authors reported in their study. The reason Arum and Roska found that so many (not so few) students improved significantly is that they computed the wrong significance test.

This particular problem is further highlighted in a paper published by the CLA itself titled *The Collegiate Learning Assessment: Facts and Fantasies*, in which they make clear that “The CLA focuses on the institution (rather than the student) as the unit of analysis . . . [and] The CLA itself does not identify the reasons why a school’s students do better or worse than expected” (Klein, et al. 3).

But for those of us not statistically inclined, there are other glaring problems with claims that this standardized test of writing can be used to measure change in student ability over time. In fact, these problems seem to echo the very same ones that caused the National Assessment of Educational Progress to question the validity and reliability of their long-term trend assessments in writing and, ultimately, to declare them not reliable or valid enough upon which to make claims about change in writing achievement over time. The first problem is whether or not the writing tasks and the measurement tools used at two different intervals were controlled to a level that would allow for valid and reliable comparison of change over time. It is important to emphasize that they *seem* to echo these problems because the authors of *Academically Adrift* will not release the actual pre- and post-writing prompts used in their research so that those who specialize in writing assessment and test development can measure the validity and reliability of their claims. This unwillingness to engage in full peer review,

especially to a degree that would allow others to determine the validity and reliability of their results through means such as replicability, certainly calls their research and motives into question.

The second problem concerns the writing tasks themselves. CLA and the authors of *Academically Adrift* emphasize numerous times that their performance-based assessments of writing are authentic and based on *general skills* as opposed to *specific content knowledge* gained through exposure to the primary texts in one's major or discipline. They point to the following performance-based assessment as representative of a task requiring only *general skills*:

The “DynaTech” performance task asks students to generate a memo advising an employer about the desirability of purchasing a type of airplane that has recently crashed. Students are informed: “You are the assistant to Pat Williams, the president of DynaTech, a company that makes precision electronic instruments and navigational equipment. Sally Evans, a member of DynaTech’s sales force, recommended that DynaTech buy a small private plane (a SwiftAir 235) that she and other members of the sales force could use to visit customers. Pat was about to approve the purchase when there was an accident involving a SwiftAir 235.” Students are provided with the following set of documents for this activity: newspaper articles about the accident, a federal accident report on in-flight breakups in single engine planes, Pat Williams’ e-mail to her assistant and Sally Evans’ e-mail to Pat Williams, charts on SwiftAir’s performance characteristics, an article from *Amateur Pilot* magazine comparing the SwiftAir 235 to similar planes, and pictures and descriptions of SwiftAir models 180 and 235. Students are then instructed to “prepare a memo that addresses several questions, including what data support or refute the claim that the type of wing on the SwiftAir 235 leads to more in-flight breakups, what other factors might have contributed to the accident and should be taken in account, and your overall recommendation about whether or not DynaTech should purchase the plane. (*Academically Adrift*, 21–22)

Of course, there is the obvious problem of the timed nature of this task, as no one of any repute would tackle such a serious writing task in ninety minutes. Perhaps more perplexing is that it is difficult at best to understand how a prompt requiring knowledge of a discipline-specific genre, a formal business

memo, about a discipline-specific subject, aerospace engineering, within the context of another specific field, risk management, could be considered a test of general knowledge. Further, this is an “authentic” test for a very, very small subset of our society—those in the position to make high-level risk management decisions. Again, it is troubling that neither the authors of *Academically Adrift* nor CLA will release the actual performance prompts used. But, if the above performance prompt is representative, as the authors claim, then it is very likely that students were not tested on general knowledge but rather very genre- and discipline-specific knowledge and, further that the genres and disciplines were different in the pre- and post-tests. Nonetheless, this ill-conceived study continues to be used as one of the primary arguments for enacting greater systems of accountability in higher education writing classrooms.

DIRECTING FUTURE ATTENTION

Margaret Strain concludes in her article “In Defense of a Nation: The National Defense of Education Act, Project English, and the Origins of Empirical Research in Composition”: “By seeing historical events as a dynamic interplay of resistance and persuasion among groups of varied power, we are able to recognize and appreciate the competing interests that inform a historical moment” (533). We would add that this type of work also allows us to chart a path forward as emerging entanglements in the struggle to control literacy are revealed. As we bring this phase of our investigation to a close, we move toward understanding how all of this may reshape composition classrooms. Specifically, we are concerned about the possible effects of the Common Core State Standards, not in and of themselves, but in and of their relationship to standardized tests of writing on the field of rhetoric and composition. The CCSS are self-described as:

a set of high-quality academic standards in mathematics and English language arts/literacy (ELA). These learning goals outline what a student should know and be able to do at the end of each grade. The standards were created to ensure that all students graduate from high school with the skills and knowledge necessary to succeed in college, career, and life, regardless of where they live. Forty-four states, the District of Columbia, four territories, and the Department of Defense Education Activity (DoDEA) have voluntarily adopted and are moving forward with the Common Core.⁷

Much like the earlier rhetoric of crisis following Sputnik that led to the National Assessment of Educational Progress and was echoed in *A Test of Leadership* and

its attendant calls for systems of accountability, the CCSS are being propelled by a fear that the United States is falling dangerously behind other countries in global tests of academic achievement. As the October 7, 2013, issue of *Time* proclaimed: “What’s driving the core standards conversation now is the ambition to succeed in a global economy and the anxiety that American students are failing to do so” (Meacham 44). This crisis rhetoric can be found in the Council on Foreign Relations Task Force’s report *US Education Reform and National Security* that argues a failing U.S. education system threatens our national security in five specific ways: “threats to economic growth and competitiveness, U.S. physical safety, intellectual property, U.S. global awareness, and U.S. unity and cohesion” (qtd. in Klein and Rice 7). Further, while critiques of the CCSS abound, overall their adoption has been swift and ongoing as textbooks are realigned, tests developed, school district rubrics restructured, and teachers trained. In fact, as mentioned in our introduction, when a small number of governors began to publicly denounce CCSS after previously adopting the standards, the group Higher Ed for Higher Standards was formed and includes over 200 presidents, chancellors, state officials, and organizations such as the American Association of Colleges and Universities (AAC&U). Much like Harvard in the 1800s, this group is working to establish processes of articulation, this time via CCSS. Perhaps not surprisingly, this coalition is part of the Collaborative for Student Success, funded in large part by the Bill and Melinda Gates Foundation (Mangan), the primary investor in the CCSS itself.

The conflicts of interest in terms of how the CCSS are being funded and implemented forebode systems of accountability and measurement that will rest heavily on writing instruction at the college level. Thomas Newkirk begins to unravel these conflicts in “Speaking Back to the Common Core”:

The Common Core State Standards are joined at the hip to standardized tests, not surprising because both the College Board and the ACT have had such a big role in their creation. It was clear from their conception that they would play a large part in teaching evaluation, a requirement for applications for Race to the Top funds and exemptions from No Child Left Behind. (4)

For example, David Coleman, who became the president of College Board in 2012, and thus overseer of the SAT, is not only one of the major initiators of the CCSS, but one of the people who convinced Bill and Melinda Gates to fund them. Bill Gates did more than simply fund their development; he “was de facto organizer, providing money and structure for states to work together on common standards in a way that avoided the usual collision between states’

rights and national interests that had undercut every previous effort” (Layton). Coleman went on to write much of the standards for math and literacy. Most recently, in many well-publicized events, he announced that the SAT will be redesigned to align with the CCSS. One of the changes includes making the essay part of the exam optional. The entanglements don’t end here. As reported in the November 3, 2013, issue of the *Chronicle of Higher Education*, the Bill and Melinda Gates Foundation hired Richard Arum, one of the authors of *Academically Adrift*, as a senior fellow on educational quality.

The influence of private foundations reaches far beyond investment in the development of the standards. For example, the National Writing Project is now significantly funded in part by the Bill and Melinda Gates Foundation, and this funding reaches down into local sites specifically in an increased effort to gain compliance with the CCSS. In 2010 The National Writing Project received a \$550,000 grant from the Bill & Melinda Gates Foundation and teams of teachers were expected to “create a model for classroom teachers in writing instruction across the curriculum that will support students to achieve the outcomes of the Common Core Standards” (“To Create”). In 2011 the Bill and Melinda Gates Foundation awarded \$3,095,593 in grant money to local sites of the National Writing Project to “create curricula models for classroom teachers in writing instruction that will support students to achieve the outcomes of the newly state-adopted Common Core Standards” (“Denver Writing Project”). In 2014, the Bill and Melinda Gates Foundation funded the Assignments Matter program. These grants were designed to “introduce large numbers of teachers to the Literacy Design Collaborative (LDC) and its tools for making and sharing writing assignments. Specifically, we will introduce teachers to the LDC task bank and jurying rubric, tools meant to support teachers in creating clear and meaningful writing prompts” (“Assignments Matter”).

While the official website for the Common Core State Standards emphasizes the flexibility teachers have in developing curriculum, the Literacy Design Collaborative belies what may appear to be support for teacher agency. In 2013, the Bill and Melinda Gates Foundation directed \$12,000,000 to “incubate an anchor Literacy Design Collaborative (LDC) organization to further expand reach and impact [of the Common Core State Standards]” (Literacy Design Collaborative, Inc.). On their official website, the LDC purports to put “educators in the lead” but only in so much as they operate within the relatively narrow parameters of rubrics designed and approved by the Collaborative. For example:

[LDC] has created a process to validate the CCRS alignment of LDC-created content. The SCALE-created “jurying” process looks at how richly the tasks and modules engage

academic content and build CCRS-aligned skills. Jurying can provide guidance on how to improve each module and is used to identify modules that are ready to share, as well as to spotlight those that reach the standards for “exemplary” that are in the LDC Curriculum Library. (“Overview”)

Furthermore, teachers are expected to use the LDC developed rubrics when assessing student work:

After a module’s instructional plan is taught and students’ final products (their responses to the teaching task) are collected, teachers score the work using LDC rubrics that are focused on key CCRS-aligned features as well as on the disciplinary knowledge shown in each piece. Visit the Rubric page for more information. (“Overview”)

The LDC claims to have “enabled” tens of thousands of teachers to prepare students for the 21st Century workforce. With a \$12,000,000 initial investment by the Bill and Melinda Gates Foundation, the LDC has the resources needed to incentivize and build professional development activities that are highly regulated and closely aligned with CCSS.

Perhaps of more direct importance to the field of rhetoric and composition is the Core to College initiative. Core to College is a sponsored project of Rockefeller Philanthropy Advisors and is funded by the Lumina Foundation, the William and Flora Hewlett Foundation, the Bill and Melinda Gates Foundation, and the Carnegie Corporation of New York. According to the Rockefeller Philanthropy Advisors, eleven states—Colorado, Florida, Hawaii, Indiana, Kentucky, Louisiana, Massachusetts, North Carolina, Oregon, Tennessee and Washington—have been provided with funds to use the CCSS to drive curricular alignment in academic courses and sequences, data and accountability, and teacher development (Rockefeller). Each of these twelve states has an Alignment Director (AD) whose job is to oversee the Core to College initiative in his or her state. WestEd has been retained to track progress of this initiative. In 2013, WestEd released their report “Implementing the Common Core State Standards: Articulating Course Sequences Across K–12 and Higher Education Systems” (Finkelstein, et. al). Interestingly, even though the primary goal of this initiative is to align course sequencing and instruction across K–12 and higher education systems, and even though there is widespread belief in the importance of course sequencing among the ADs, the report concludes “the CCSS do not appear to figure prominently into states’ current course sequencing discussions” (29). In their related report “Exploring the Use of Multiple Measures for Placement into College-Level Courses,” released in 2014 and based on the a survey of ADs,

WestEd affirms research evidencing that standardized tests alone are not the best means for determining college admissions and placement (Bracco et. al). This is important given the research we previously detailed on the use of standardized tests for this purpose. The report discusses the range of measures in Core to College states that are being considered for college placement. Perhaps all we can take away from the WestEd studies of Core to College is that the effectiveness of Common Core State Standards in creating greater alignment and collaboration among K–12 and higher education is quite mixed. The mixed results of the Core to College initiative make it difficult to determine ongoing effects of this type of work. The Core to College initiative formally ended in 2014, although some states are certainly continuing this work and it will be important to see if it will lead to lasting and impactful K–12 and college collaborations. While we might be optimistic about the rich opportunities K–12 and college collaborations can yield, given how these efforts are being funded and how often they are used to establish ever greater systems of accountability and control over our K–12 classrooms, we must be cautious and critical optimists as we move forward.

All of this raises questions about who is driving U.S. higher education these days. Of course, higher education in the United States has always been shaped by multiple competing forces. For example, beginning in 1938 with Earnest Hollis’ book *Philanthropy Foundations and Higher Education*, many researchers have documented the influence that private foundations have had on reforming higher education. In a study published in 2011 by Cassie Hall—using a review of academic literature, an analysis of public discourse from a wide variety of media, ten years of secondary data on philanthropic giving to higher education, and interviews with five senior-level professionals—Hall shows that there has been a fundamental shift in the relationship between foundations, higher education, and the control of public policy. Historically, foundations shaped higher education primarily through direct incentives to institutions with a focus on capital construction, academic research or programmatic efforts (Hall 16). But as Hall demonstrates in her analysis of the changing relationship between foundations and higher education, “recent foundation behavior suggests that a new approach to higher education philanthropy has emerged over the past decade, one that emphasizes broad-scale reform initiatives and systemic change through focused, hands-on public policy work” (2). This new approach to foundation work is being referred to as “advocacy philanthropy.” Hall argues that foundations’ “overt focus on public policy advocacy within specific state and local contexts will have a significant impact on higher education in the United States” (50).

As a conclusion to her study, Hall discusses the possible benefits, concerns, and emerging outcomes of this shift. Potential benefits of advocacy philanthropy

include the attention foundations are drawing to important problems; creating a sense of urgency in the search for solutions; the effectiveness of grantmaking in bringing key actors together; and the ability of foundations to scale up reforms to achieve substantive change (96–100). Among the concerns are foundations’ lack of external accountability and their concentration of power away from practice; the potential of their large-scale prescriptive grants to stifle innovation; and the extensive, perhaps excessive, influence gained by foundations through such advocacy (96–100). Emerging outcomes also raise issues to consider, such as diminishing funds available for field-initiated academic research, a shift from local focus to a national one that could affect changes to higher education power structures, and the lessening of trust in higher education institutions (83–92). Hall concludes, “the Bill and Melinda Gates Foundation and the Lumina Foundation for Education have taken up a set of methods—strategic grantmaking, public policy advocacy, the funding of intermediaries, and collaboration with government—that illustrate their direct and unapologetic desire to influence policy and practice in numerous higher education arenas” (109).

One of the areas in which Hall’s concerns are most apparent is in how the Bill and Melinda Gates Foundation is funding the CCSS—our nation’s first set of national standards marking perhaps one of the biggest public policy shifts in education to date. The analysis by Hall evidences that “college ready funding has been the largest funding priority for the Gates Foundation” (14). And when we refer to funding we are not talking about just the research, design, and implementation of the Standards, but also how they are funding support networks. For example, providing financial support to sites of the National Writing Project that agree to teach teachers how to meet CCSS, partnering with other foundations to support the Core to College initiative, founding the Collaborative for Student Success with other foundations whose sole purpose is to market the CCSS, and, more recently, funding Higher Ed for Higher Standards—a project of the Collaborative for Student Success designed to show that the CCSS are backed by our higher education leaders.

Understanding the role of accountability is crucial to the cautious and critically optimistic stance we take toward the CCSS and Core to College. Recognizing the competing forces at play, we see opportunity in the fact that for the first time, national standards have been established that attempt to put writing (and reading) on equal footing with science and math. We position ourselves as critical optimists because we believe the CCSS, while flawed, have value. Further, we hope that initiatives such as Core to College will lead to greater collaboration between high school and college faculty. However, as our historical sketch exhibits, at no other time have so many competing interests exerted such powerful and far-reaching force on U.S. classrooms in the name of accountability. And, our

continued over-reliance on standardized testing is not only alarming, but also not producing the intended effects. As we chart a path forward, our next step must surely be to create opportunities to firmly establish student and teacher agency in the research, practice, and assessment of writing so that we can acknowledge the changes that need to be made to education without succumbing to the siren's call of crisis.

NOTES

1. As evidence of this increase we can look to the Collegiate Learning Assessment (CLA), which has grown the number of participating higher education institutions to 700 from its inception in 2002. Further, CAE (Council for Aid to Education), the organization that administers the CLA, is working with those developing Common Core State Standards Assessments to ensure alignment between their standardized tests and those used at the college level such as the CLA (Council for Aid to Education).
2. Throughout this book we focus on PARCC, but there is another consortium that has also developed CCSS aligned standardized tests—The Smarter Balanced Assessment Consortium. Because we don't intend this book to focus primarily on an analysis of these consortia, we chose to focus on PARCC as just one example of the current state of standardized testing in relationship to high school and college curricula both because it is the more controversial of the two consortia and because we both happen to live in PARCC member states.
3. Many historians agree that the first standardized tests to include writing were administered in China as early as 1115 A.D. These were known as "Imperial Examinations" and covered the Six Arts: music, math, writing, knowledge of the rituals of public and private life, archery, and horsemanship (Ward 44). The Imperial Examination was essentially a civil service exam that was open to nearly all males "and became the most important avenue to position, power, and prestige in China" (Hanson, 186).
4. For more on the role of the National Defense of Education Act on the shape of rhetoric and composition as a field, see Margaret Strain's "In Defense of a Nation: The National Defense Education Act, Project English, and the Origins of Empirical Research in Composition."
5. For more information, see <http://www.learningoutcomesassessment.org/AboutUs.html>.
6. For more information, see <http://digitalcommons.unl.edu/peerreviewteaching/>.
7. For a fuller discussion of the CCSS, see <http://www.corestandards.org/about-the-standards>.