Preface

THIS BOOK HAS BEEN GESTATING for almost twenty years. It was conceived, unbeknownst to me at the time, when a program officer at the National Science Foundation asked if I thought that a collegiate version of NAEP could be built.[1] I wondered why the government would want a one-size-fits-all, largely multiple-choice test for all colleges and universities in their full diversity. What good might come of information provided by a collegiate NAEP with scores reported publically in league tables? Why adopt wholesale for higher education an assessment built to monitor mandatory precollegiate education?

I paused then and said that that wasn't a good idea, and, if it was tried, I would oppose it. I didn't see how a single, narrowly gauged achievement test of basic skills could be developed in a manner sensitive to the diversity of education and missions in the nation's institutions of higher education, including the development of higher-order cognitive abilities and personal and interpersonal skills. I didn't see how information provided by a single, general test could be used to improve teaching and learning in higher education. And I didn't see why it would be appropriate to adopt a solution to mandatory precollegiate education for elective higher education, knowing the strengths and limitations of large-scale assessments in an accountability context, as well as the political uses and misuses that have been made of such tests.

I then lost sight of the question of higher-education accountability for a couple of years until a friend, a music professor at a small midwestern liberal arts college, phoned. He had been appointed to a campus-wide committee charged with responding to the North Central Accreditation and School Improvement Association's mandate to assess student learning. He wondered if I thought it appropriate that his college replace its current system of assessing students with an on-demand, multiple-choice test of largely factual and procedural knowledge in the humanities, social sciences, and sciences to meet accreditation demands. He explained that currently all seniors completed a capstone course with high performance expectations; for example, his opera students had to stage an opera, among other requirements. This, he thought, was more relevant to his students' achievement than a humanities multiple-choice test. He asked if I saw something wrong in his thinking. I told him that I didn't think so and suggested that perhaps his committee and his college were overreacting to the accreditation mandate. The questions raised about a collegiate NAEP returned in a new context.

A few years later, learning assessment and accountability came to my attention again, this time in a newspaper article. On Sunday, September 27, 1998, the *New York Times* alerted readers to the New York State Education Department's plan to evaluate public and private colleges and publish the findings as early as 2001. The department planned to convene a higher-education advisory council of college presidents to guide its efforts to produce a "report card" based on a mandatory test for the state's higher-education institutions, public and private. New York was following a trend in the United States (and other countries, such as Britain and Australia) toward increased higher-education accountability. The State University of New York, for one, demurred; the proposal needed further study; a system-wide committee was appointed to do the review.

The New York situation weighed on me. What alternatives were there to one-size-fits-all assessment? What alternatives were there to U.S.-style accountability? Is the K-12 vision embodied in the No Child Left Behind federal legislation the only reasonable option?

These questions were on my mind when a program officer from the Atlantic Philanthropic Service Company (APS), Myra Strober, invited me to lunch to talk about trends in higher education, especially the push for accountability. Myra had just taken a leave from Stanford to direct APS's higher-education grants program and was in the process of framing a portfolio of new projects. When I told her my concerns about accountability trends, she, too, became concerned about the possible unintended negative consequences for higher education.

My discussion with Myra ultimately led to support for the work contained herein, in large part a grant from APS (now called Atlantic Philanthropies). Once Myra asked for a proposal, she turned everything over to Jim Spencer, her predecessor, to avoid any conflict of interest, she and I both being from Stanford.

In this text I examine current practice in assessment of learning and higher-education accountability. By “assessment of learning” I mean the use of both direct measures of achievement (e.g., certification examinations) and ability (e.g., Graduate Record Examination, Collegiate Learning Assessment) and indirect measures (graduation and retention rates, time to degree, job placement and employer satisfaction, and student surveys of engagement). By “accountability” I mean the collection, provision, and interpretation of information on higher-education quality sought by educators and policy makers who have responsibility for assuring the public and “clients”—students, parents, businesses, and government—that invest in education, training, or research.

The goal of this text is to provide education policy makers—in the academy, in government, and in the public—with an overview and critical analysis of options for crafting learning assessment and accountability systems that meet needs for campus teaching and learning improvement and external accountability. Along the way, I identify alternative conceptions of and procedures for assessment and accountability systems, some of which may substantively improve college teaching and learning, both in general education and in the disciplines, while at the same time informing external audiences.

The book begins by introducing the higher-education policy context in the United States and the current demand for learning assessment and external accountability (Chapter 1). A number of tensions emerge, not the least of which is between the formative (institutional improvement) and summative (comparative) functions of accountability and who controls that agenda. A second, related tension is whether and to what extent campuses’ performances are publicly compared with one another.

Chapters 2 through 5 address the quest to assess student learning. Chapter 2 distinguishes among direct and indirect measures of learning, arguing that indirect measures do not measure learning, and distinguishes learning (relatively permanent change in behavior over time) from achievement (level of academic performance at one time point) and propensity to learn (level of achievement within a student’s reach with minimal scaffolding). A framework is then presented for considering assessment of learning and achievement, ranging from knowledge and reasoning within a domain (e.g., quadratic equations) or major (e.g., mathematics) to broad reasoning, decision making, and communicating within the sciences, social sciences, and humanities; to quantitative, verbal, and spatial reasoning; to general ability. The framework locates current learning assessments and provides a crosswalk among different notions and recommendations for measuring learning outcomes.

In Chapter 3, the 100-year history of learning assessment in higher education is sketched, drawing lessons to be learned from the past for the design of learning assessment and showing that the current debate is not new. I then turn to currently available, externally provided learning assessments and what they attempt to do, concluding that the recent Collegiate Learning Assessment (CLA) offers a great deal of promise. Chapter 4 provides detailed information about the CLA, as it is, arguably, the newest, most innovative assessment of college learning today and relatively little is known about its philosophy and technical qualities.

The last chapter of the learning assessment sequence (Chapter 5) examines undergraduate learning assessment as practiced on campuses; campus-based assessment efforts are essential to meet both formative and summative accountability demands. External assessments signal areas in need of improvement by benchmarking campus performance against the performance of campuses viewed as peers; local campus information is needed to pinpoint challenges and to conjecture and test out possible ways of improving learning. The variability among even exemplary campus assessment programs becomes immediately apparent—in how they were started and are sustained, in what they did (do), and in their intended and unintended consequences for student learning and teaching. The goal here is to identify programs and their implementation and operation that appear to have salutary effects on teaching and learning and draw lessons for the design of learning assessment and accountability systems.

Chapters 6 through 9 focus on accountability. Chapter 6 addresses the centrality of information in accountability and the "cultural conflict" among academe, government, and clients. Although conflict is inevitable, it can nevertheless be productive when cool heads prevail; given the politics of higher-education accountability, the assumption of cool heads is tenuous. Rather, there is considerable room for mischief on all sides.

Chapter 7 examines the role of accountability in a democracy, drawing implications for the role of accountability in higher education. There is a tension between accountability for formative (improvement) and summative (external informative) purposes, as well as between accounting for actions and accounting for outcomes. Moreover, the application of accountability to higher education gives rise to issues such as the presumption of control and causality, the role of sanctions, and the power of whoever controls the stories or accounts that that provide interpretations of accountability information for the public. What becomes clear is that accountability is a powerful policy instrument but a delicate one, one that, if misapplied, may lead to as much mischief as good.

Chapter 8 explores current state-level accountability practices in the United States. How many states have such practices? What do these practices look like? How do they vary? What consequences, intended and unintended, do they appear to have? Performance reporting of some kind dominates in a wide variety of forms. While myriad indicators are published, few states actually report direct measures of learning. And states report so many indicators that performance reports lack focus; the public and policy makers are overwhelmed by data.

In Chapter 9 I analyze accountability systems in different parts of the world, including the European Community generally, especially England and Scandinavia, and Australia, New Zealand, and Hong Kong. Clear alternatives to current practice in the United States (although that is changing) exist. Outside the United States, quality assurance has taken hold. Accreditation, assessment of learning for cognitive and responsibility outcomes, and quality assurance are, in a certain combination, shown to be viable alternatives to current practice in the United States.

The book concludes (in Chapter 10) by setting forth a vision of an assessment and accountability "system" that integrates the findings from the previous chapters. I envision a multifaceted approach to the assessment of learning that includes cognitive outcomes in the majors and in broad abilities, including critical thinking, analytic reasoning, problem solving, and communicating. This vision of learning assessment also encompasses individual and social responsibility outcomes, including the development of personal identity, emotional competence, resilience, and perspective taking (interpersonal, moral, and civic). Learning assessment, both internal and external to colleges and universities, is a centerpiece for a quality assurance system of accountability that incorporates accreditation and assessment. Such a system provides both formative and summative information to higher educators, policy makers, clients, and the public while addressing the tension of conflicting policy and education cultures.

Inevitably, some readers will find some topics of little or no interest. For example, I have not distinguished between public and private four-year institutions or distinguished institutions by Carnegie classification. I believe that because I have been a faculty member and dean at both public (University of California at Los Angeles and University of California at Santa Barbara) and private (Stanford) universities, what I say here can be applied fruitfully across these institutional types (although perhaps more to some types, such as liberal arts colleges, than to, say, research universities). Setting clear goals, building programs to reach them, monitoring progress, and feeding back findings that provide a basis for improvement and experimentation would seem to be beneficial across the spectrum.

Moreover, community colleges and for-profit institutions are not addressed specifically. To be sure, what is said about four-year public and private colleges and universities here may be informative for community colleges and for-profit institutions. However, none of the examples or case studies presented draw on these institutions. Nor was consideration given to their differences and what might be said about them that would differ from what is said about four-year campuses. Simply put, they were beyond the scope of this work.

I am indebted to many colleagues, not the least of whom were program officers at APS overseeing this work—Myra Strober, Jim Spencer, Ted Hullar, and Ray Handlan. I have already described Myra's role. Jim Spencer, in his review of my proposal, said it all sounded academic and why didn't I immerse myself in practice? (Jim's an engineer.) His advice led to my involvement in the creation of the Collegiate Learning Assessment. The responsibility for guiding my grant, however, largely fell on the shoulders of Ted Hullar, who replaced Myra as higher-education program director for APS. His strong support for the project and his patience in the face of slow progress were motivating and greatly appreciated. Ultimately, as the APS higher-education program was phased out, Ted saw to it that I had the resources needed to complete the work and write this text; Ray Handlan did the same, following Ted as my contact.

I am also deeply indebted to Maria Araceli Ruiz-Primo, formerly of Stanford University and now at the University of Colorado–Denver. She helped design, analyze, and report the empirical research conducted for the book. And she patiently read and critiqued a number of chapters. I am also indebted to Blake Naughton, who, as a graduate student at Stanford, helped conceive and design the study of state accountability systems; to Anita Suen, who assisted with research reported in Chapter 8; and to Gayle Christensen, now at the University of Pennsylvania, who as a graduate student and then a Humboldt Fellow at the Max Planck Institute in Berlin provided research support for the chapter on international approaches to accountability (Chapter 9). Finally, a debt of gratitude goes to Lee Shulman, who provided support, advice, wisdom, and encouragement throughout the project.

My colleagues at the Council for Aid to Education—Roger Benjamin, Roger Bolus, and Steve Klein—provided invaluable support for the chapter on the Collegiate Learning Assessment (Chapter 4). My experiences with them in the development and now the use of the CLA proved formative in my thinking about the assessment of learning and its role in higher-education accountability.