A Measure of Failure

> The idea underlying the implementation of written examinations [in the nineteenth century], that they could provide information about student learning, was born in the minds of individuals already convinced that education was substandard in quality. This sequence—perception of failure followed by the collection of data designed to document failure . . . offers early evidence of what has become a tradition of school reform and a truism of student testing: tests are often administered not just to discover how well schools or kids are doing, but rather to obtain external *confirmation*—validation—of the hypothesis that they are not doing well at all. (Office of Technology Assessment 1992, 108)

I had marked "Nota Bene" on this passage more than ten years ago as a graduate student. It struck me as a very important yet underexplored observation. Why this evident preoccupation with the failure of public education? I became particularly interested in the significance of the present trend to use the results of standardized tests to justify the elimination of public education "as we know it" when I began to understand that this technology originated to justify and help erect our public education system.[1]

This observation led me to carry out the case studies of Horace Mann and Alfred Binet that are the mainstay of this book. In each, the documentation of failure—the failure of the then existing Boston public schools, the extensive "retardation" existing in French schools and their failure to properly identify and educate "mental defectives"—facilitated reforms of those time periods. Advances in testing accompanied the documentation of failure, where mental tests were said to not only identify but also remedy the situation. Thus, standardized tests originated in a crucible of failure. In each case, the reader will see that standardized tests of either achievement

or ability were tools used to institute and justify substantive changes in the governance and functioning of education.

This book argues that standardized testing technology originated as a tool of liberal (or representative) democracy that enables the system to present as egalitarian a social order that values social difference. In this context, the once unimportant idea of mental ability became a basis for political rights (Carson 1994). As public schools developed, they increasingly assumed the role of marking such abilities. Academic tests emerged to serve this purpose constituting an ideological tool for justifying social inequality (fair competition), a political tool for securing state power and affirming the power of public office and the professions (accountability), and a philosophical tool for justifying public governance of education and for educating "the masses" or public (reason or judgment).

The case studies show that tests of achievement and ability emerged to mark the then existing educational institutions as failures, for these existing arrangements could not meet the needs of the then emerging social orders. The new tests served as instantiations of the new model and purpose of education established by a new authority or governing arrangement. An emphasis on testing is, this research suggests, an outgrowth of a crisis evident in an intensifying political fight between factions of elite, on the one hand, and between social classes, on the other. This book exists to argue out the aforementioned to provide background to the present testing movement. It should make clear that neither in the past nor the present is testing mainly about "improving education." It is, instead, about control over the purpose and nature of schooling.

Failure and Crisis

The emphasis on failure in reports about public education is as significant as it is problematic. Debates about public school failure tend to give a great deal of emphasis to interpreting standardized test data, with the future of public education purportedly hinged to these data, especially when discussed within the framework of the No Child Left Behind Act (NCLB) and its mandate that public schools come under the discipline of the federally subsidized "free market" when Adequate Yearly Progress is not reached (e.g., mandates for charters, and the for-profit provisions of tutoring services). Methodological questions as they relate to conclusions about the efficacy of a particular program or public education as a whole have long dominated the academic discussion (e.g., Berliner 1993; Bracey 2002; Saltman 2005; Stedman 1996). But lacking in these debates is an understanding of the conceptions of failure and of crisis and the significance of such designations.

Foremost is the need to acknowledge that failure points to the problem of reproduction: the claim that an institution is failing is a claim that it cannot reproduce or serve its social function. It is the claim that something cannot be repaired, but must instead be replaced. But failure is also a designation of value. To designate something as a failure is to devalue it; the discourse of failure serves to discredit. Failure, too, is real—even propaganda about failure has to have a real basis. The fact is, public schools have failed to meet the educational needs of large sections of the people—what is unique in the present is that public schools appear to be unable to play their social reproduction function as in the past. The accountability movement and the central role standardized tests play within that movement is in part an outgrowth of this reality (see Dorn 2007).

Generally speaking, failure means "omitting to perform something due or required," a notion that implies a standard, an expectation of performance, especially when performance is understood as role or responsibility. Importantly, the Oxford English Dictionary (OED) gives another sense of failure as "becoming exhausted or running short, breaking down in health, declining in strength or activity." Thus failure can signify breakdowns in social relations or institutions, which may be "breaking down in health, declining in strength," so to speak.[2]

Originating in the seventeenth century, the word *failure* referenced a breach in justice and the problem of succession (bloodlines). Immediately, then, one sees its original connection to politics. Justice, a notion connected to legal systems and the state, is required for the legitimacy and thus longevity of a political order. Outlining the development of the conception of failure, the OED gives this example from Mountstuart Elphinstone's 1815 *An Account of the Kingdom of Caubul*: "On the failure of issue . . . an adopted son succeeds." This suggests in particular that failure references a breakdown in the reproduction of political systems. Reformers such as Mann and Binet, the reader will see, sought to resolve social contradictions by giving a greater role to government-controlled schools in reproducing the social order. Standardized exams were developed as markers of failure, and stood as justifications for and symbols of the changes reformers sought.

A time period marked by failure is a turning point, where what has been is dying away, no longer sustainable by any social force, and what is coming into being has yet to take hold, with no force clearly in control. These turning points are by definition periods of crisis, and I take the two case studies in this book to be periods of crisis. The word *crisis*, according to the OED, comes from the Greek meaning to separate, to decide; crisis is given as the point in time "when it is decided whether an affair or course of action shall proceed, be modified, or terminated," or more generally an "unstable state of affairs in which a decisive change is impending."

Thus, instead of assuming a "manufactured" crisis—a position all too common in my view, one that tends to presume that society develops according to the wishes of those in elite positions—I take assertions of crisis seriously. It has been said that such proclamations are mere hyperbole, the rhetoric of clever politicians with a hidden agenda. This is, of course, partly true and generally well documented (Bracey 2002; Saltman 2005). But this is limited in terms of analysis.

Rendering the crisis in education and society as a fabrication of the elite veils why the elite (the bankers, industrialists, and key politicians) would work so hard to destroy that which those in their position one hundred fifty years ago worked so hard to build as necessary for maintaining their political system and their power? On what basis is this institution—public education—no longer serving its functions, from their point of view? Why use tests to discredit this institution when they proved so useful in establishing it? And what significance and opening does this crisis have for defending rights and for people's genuine empowerment and public participation?

A final word is in order regarding the role of standards in discussions of failure and crisis. The issue is not so much to debate whether schools are failing, for this is, in fact, an abstraction. "Failure to perform a duty or action" is a specific thing, that is, it refers to failure to perform something. The thesis is that the standards used to judge the success of schools have changed, and that this change in standards is about shifts in power and purpose, not "school improvement." What schools are expected to do by those officials who now wield power is different from the past, and this difference is reflected in their adoption of different standards than those put into place by their predecessors. That business leaders were, for example, directed by the Business Roundtable to sit on "cut score committees" to ensure high levels of failure is a remarkable example of the role of standards in establishing power (Business Roundtable 1998) and more generally of standardized testing as a measure of failure.

Chapter Overview

Oriented by historical sociology, chapter 2 presents a heuristic for analyzing standards based in a review of literature tracing the nature and function of standards of both physical (e.g., length) and extraphysical (e.g., academic achievement) phenomena. Standards are explored as weapons in the quest for power, as a means to give material expression to a philosophy or aim, and as embodiments of the social values of a culture or class. The heuristic poses four interrelated themes to guide the analysis. Chapter 3 explores the notions of achievement and ability as forms of vertical classification, where notions of ability and achievement are explored as basic tenets of

Anglo-American political thought, that natural distinctions should be the basis for ordering society. Chapter 4 argues that standardized tests—as concrete expressions of the cultural meaning of achievement and ability—function to assess social value. Social value is the idea that human beings exist in distinct hierarchical groupings in terms of their value or worth. It also refers to the process by which individual or group value is socially attached to a position within a hierarchically structured social system. Standardized testing—or the theory and practice known as "psychometrics"—the fourth chapter argues, is not a form of measurement. Psychometrics is best understood as the development of tools for vertical classification and the production of social value.

Chapter 5 presents and examines the political origin of standardized testing in the emergence of nation states and political arrangements favoring capital as a class (against feudal political arrangements) and the role their new theory of governance gave to reason and the so-called natural aristocracy. As it is well suited to "develop new theories capable of providing more convincing and comprehensive explanations for historical patterns and structures" by focusing in on the growth of national states (Bonnell 1980, 161; see also Green 1990; Tilly 1981), the approach of historical sociology is adopted. The role given to formal education in nation building is explored as an important condition for the emergence of standardized tests of achievement and ability and the linkage of mental ability to rights, especially the right to govern and the role given to enlightened public opinion in governance.

Adopting Bonnell's (1980) method of case-illustration, whereby different cases are compared to a single theory or concept, two key points or cases in the development of standardized testing in education are analyzed. The first case, involving Horace Mann, is explored in chapter 6. It examines the Boston Grammar School Committee's efforts to develop what is considered the first large-scale use of standardized test methods to "secure positive, objective information about the products of classroom instruction" (Caldwell and Courtis 1925; Resnick 1982). While for hundreds of years various oral examinations and school visitations had been used to assess educational progress and select and certify pupils, this development marks the beginning of achievement testing as we know it today. Mann's use of these tests helped bring the Boston public schools under state supervision, establish practices consistent with notions of equal opportunity in education while also supporting the common school reform agenda, including everything from the age-graded classroom to state-run teachers colleges.

After decades of unfruitful work to measure mental capacities of various kinds, and wide variation in the definition and classification of intelligence, chapter 7 explores the work of Alfred Binet and his colleagues to

develop the first practical means to measure "intelligence." Since that time, this "IQ" test has become the standard by which intelligence in children and adults is measured and defined (Sarason 1976). Binet's work took place in the context of the secularization of the French school system. It helped establish common standards for the identification of students who would not benefit from "normal instruction." This work was instrumental in institutionalizing "tracking" and a new model of "equal opportunity" where different groups of students were to receive different forms of education based on their future roles in society (Coleman 1977). Binet's test was guided by a desire to "engineer" society and avoid social unrest. These historical points in the development of standardized tests are examined in chapter 8 using the previously outlined heuristic. In particular, they enable us to explore the ways in which standardized tests of achievement and ability functioned to institutionalize notions about the role of merit in establishing a legitimate political order, and the role of public education in informing that merit and validating its public expression. Both cases reveal standardized tests as mechanisms for the assessment of social value, a practice embedded in a larger political project that rewrote the rights and responsibilities of both government and citizen. The book concludes with a summary of the case studies in light of the analysis and heuristic outlined. Implications for these findings for the present context are suggested.