THE HISTORY OF THE BELL CURVE: SORTING AND THE IDEA OF NORMAL

Lynn Fendler and Irfan Muzaffar

Department of Teacher Education
Michigan State University

Abstract. Bell-curve thinking, as a model of distribution of success and failure in society, enjoys a perennial (ahistorical, objective, and law-like) status in education. As such it provides a rationale for sorting (tracking or streaming) practices in education, which has led many educators to criticize both bell-curve thinking and associated sorting practices. In this essay, Lynn Fendler and Irfan Muzaffar argue that the existing critiques of bell-curve thinking ring true for people who believe that the purpose of schooling is to promote a more equitable redistribution of resources in society; however, these arguments do not criticize the law-like character assumed for a bell curve as a representation of social reality. To extend these critiques, Fendler and Muzaffar focus on the history of the bell curve, from a representation of binomial probability, to a bearer of real things in nature, and finally to a set of expectations about how people should behave. They ultimately argue that the acceptance of bell-curve thinking in education is part of a recursive project of governance and normalization.

> Educators might say that all children can learn, but the man in the street carries a bell curve in his head.
> — Gerald Bracey, *The Final Exam*[1]

There is a widespread belief that things of the world distribute themselves according to a model depicted by the normal curve. Known popularly as the bell curve, and mathematically as the Gaussian curve, this model holds that most phenomena occur around a middle point, while few occur at either the high or low extreme ends. An assumption of bell-curve distribution permeates educational projects on several dimensions that far exceed the scope of the infamous book by Richard Herrnstein and Charles Murray.[2] For example, pass and failure rates are determined when tests are graded on a curve in which most people will receive an average grade while relatively few will excel or fail; experimental designs and standardized tests are revised until results conform to a bell curve; guidance counselors assume a normal distribution when they interpret test results; polls and survey results are represented on a curve; and many people believe that all sorts of resources — from financial to artistic — are distributed in a population according to a normal curve. The assumption of the bell-curve distribution allows for such concepts as *acceptable rate of failure* and *the Average Student*. This means that it

1. Gerald W. Bracey, *The Final Exam: A Study of the Perpetual Scrutiny of American Education* (Bloomington, Indiana: Technos Press, 1995).

2. There has been vigorous debate since the publication of Richard J. Herrnstein and Charles Murray's *The Bell Curve: Intelligence and Class Structure in American Life* (New York and London: The Free Press, 1994). The literature in response to this book is so extensive that we will not repeat the criticisms or rejoinders in this essay. For a sampling from various sides of this debate, see Steven Fraser, ed., *The Bell Curve Wars: Race, Intelligence, and the Future of America* (New York: Basic Books, 1995); Russell Jacoby and Naomi Glauberman, eds., *The Bell Curve Debate: History, Documents, Opinions* (New York: Times Books, 1995); Brian Beatty, http://www.indiana.edu/~intell/bellcurve.shtml; and Claudia Krenz, http://www.claudiax.net/bell.html.

has been made reasonable to assume that in any educational task, a few people will excel, most will be satisfactory or average, and a few will fail. Because so many people believe that the bell curve represents the way things are in nature, the ideal of a normal distribution has been naturalized in education and, to some extent, in U.S. society at large.

Sorting students by ability is one among several competing purposes of education.[3] Educators who promote sorting often justify those practices on the basis that a bell curve represents the normal distribution of things in the natural world.[4] Social Darwinism and normal distribution patterns have provided justification for norm-referenced standardized tests, grading on a curve, detection of "at risk" students, and the whole concept of an intelligence *quotient*. Insofar as the normal curve is held to be a model of natural distribution, the normal curve is regarded as an objective basis for sorting people. The bell-curve model of distribution has been taken for granted in education because it is generally accepted to be a fact of life.

On the other side, arguments against sorting take issue with the establishment of social inequalities, but they typically do not engage the terms of debate used by pro-sorting advocates. That is, while anti-sorting arguments speak of flattening the curve, the bell curve itself remains constitutive of the debate. In general the critiques of bell-curve sorting assume that schools are institutions whose practices serve to maintain the status quo of social differentiations.[5] To support their critiques, these arguments cite evidence such as the high correlation between socioeconomic status and test scores. Thus, it is implicitly assumed that if somehow society ceased to reproduce itself, the correlative relation between the test scores and socioeconomic status — and therefore the shape of the distribution curve — would change. However, this form of anti-sorting argument has an unfortunate side effect: to contest the reproduction of society, and the correlative reproduction of the bell-curve shape, this approach has to naturalize the reproduction perspective. In this essay, we do not assume that schools are *necessarily* institutions

3. Other purposes include preparing people for the workforce and educating citizens for a democratic society. See, for example, Herbert M. Kliebard, *The Struggle for the American Curriculum, 1893–1958* (New York: Routledge, 1986); David F. Labaree, "Essay Review: Politics, Markets, and the Compromised Curriculum," *Harvard Educational Review* 57, no. 4 (1987): 483; and Jeannie Oakes, *Keeping Track: How Schools Structure Inequality* (New Haven: Yale University Press, 1985).

4. F. Allan Hanson, *Testing Testing: Social Consequences of the Examined Life* (Berkeley: University of California Press, 1993); and Kliebard, *The Struggle for the American Curriculum*.

5. See, for example, Basil Bernstein, "Class and Pedagogies: Visible and Invisible," in *Education: Culture, Economy, Society,* ed. Amy Stuart Wells (New York: Oxford University Press, 1997), 29–79; and Samuel Bowles and Herbert Gintis, *Democracy and Capitalism: Property, Community, and the Contradictions of Modern Social Thought* (New York: Basic Books, 1986).

LYNN FENDLER is Associate Professor of Teacher Education at Michigan State University, 362 Erickson Hall, East Lansing, MI 48824; e-mail <fendler@msu.edu>. Her current areas of scholarship include discursive histories of disciplines, genealogical studies of subjects, and philosophies of food.

IRFAN MUZAFFAR is a Doctoral Candidate in Teacher Education at Michigan State University; e-mail <muzaffar@msu.edu>. His primary areas of scholarship are philosophy of education and the history of education.

of reproduction, although we do acknowledge that reproduction is one of many things that can happen in schools.

Our purpose here is to criticize one of the fundamental justifications of educational sorting practices, namely, the belief that a normal-curve distribution is a representation of real things in nature. Like many such assumptions that remain in place and are repeated frequently at multiple sites, the bell curve has acquired the status of an objective law. We also argue that this naturalization of the bell curve is unjust because it perpetuates the inevitability of failure; we hold that the idea of failure should be attributed to popularly ascribed features instead. But unlike other arguments against sorting that replace one kind of naturalization with another — as occurs in the case of social reproduction critiques — we use a specific kind of historical critique that undermines the durability of the bell curve's allegedly real-world basis.[6]

In this essay we critique the naturalization of the normal curve by investigating the history of the normal curve and its relation to modern educational projects. This "critical and effective history" is a substantive aspect of the critique itself, so our argument has two interwoven dimensions.[7] One dimension examines the history of the bell curve and its relation to educational theory. The other dimension is a historiographical intervention; our approach departs from the rhetorical forms that anti-sorting arguments have taken in recent U.S. literature. The purposes of this historical critique are (1) to undermine the allegedly realist basis for sorting, namely, the assumption that a bell-curve distribution is normal; and (2) to enact a critique of ahistorical objectivism in a way that does not appeal to timeless principles as premises in the argument.

In the field of education, there are certainly people who enthusiastically support sorting by means of high-stakes competition in educational testing; we are not talking to them. Rather, we are addressing more socially minded educators who resignedly accept what appear to be natural laws of unequal distribution. These are people who bemoan the injustices of sorting, but feel compelled to live in accordance with objective facts of the world. So the argument of this essay relates a history of scientific thinking in which the idea of the bell curve was transformed several times: from a model of coin tossing, to a means of minimizing error in measurements, to a representation of regularities in births and deaths, to a fabrication of the Average Man, to a depiction of patterns displayed by aggregations of people, to a set of moral laws, and finally to a standard of normal behavior in which average is the same as desirable. This overview of the history of bell-curve thinking makes it clear that sorting mechanisms are inventions that reflect historical constructs of race, sex, morality, and religion. This critique is designed so that

6. Ian Hacking calls this approach *historical ontology*: "By historical ontology I mean a description of how it is that local participation in education, as a growing imperative of the discourse on transnational human development, came into being." Ian Hacking, *Historical Ontology* (Cambridge, Massachusetts: Harvard University Press, 2002).

7. Mitchell Dean, *Critical and Effective Histories: Foucault's Methods and Historical Sociology* (New York: Routledge, 1994).

people who support educational sorting practices cannot so readily appeal to natural laws of distribution in order to justify those actions, and people who object to sorting do not have to worry that they are defying the laws of nature in their quest for social justice.

ARGUMENTS AGAINST SORTING

Assumptions about the normal curve have beset the fair-minded educators who look upon public education as a great leveler. Benjamin Bloom recognized the sanctity accorded the curve in educational discourse thirty years ago, and revolted against it:

> The normal curve is not sacred. It describes the outcome of a random process. Since education is a purposeful activity in which we seek to have the students learn what we teach, the achievement distribution should be very different from the normal curve if our instruction is effective. In fact, our educational efforts may be said to be *unsuccessful* to the extent that student achievement is normally distributed.[8]

This is a key point that underlies many tensions in educational theory. If it is assumed that our educational efforts will result in an acceptable rate of failure, then the curriculum will be designed not only to foster a populous middle ground but also to guarantee a number of failures. However, if we take Bloom's admonition to heart, it is relatively easy to see that bell-curve thinking counteracts purposeful educational activity and, by guaranteeing failure, denies *even the possibility* of success to a certain number of people.

Another interesting but typically overlooked aspect of Bloom's argument is that bell-curve thinking has negative emotional consequences for students. In fact both of Bloom's chapters in the classic collection *Mastery Learning: Theory and Practice* are primarily about students' feelings. His first chapter has four subheadings: "Interest," "Attitudes," "Self-Concept," and "Mental Health."[9] The emotional impact of failure was Bloom's primary concern. The second chapter is about the need to adapt teaching to "particular kinds of learning."[10] At the same time, Bloom assumed that the purpose of education was to prepare competent workers. Bloom's argument against bell-curve thinking and in favor of mastery learning was based on the premise that the increasing demands of the workforce required that a higher percentage of students acquire a higher level of skills:

> The complexity of the skills required by the workforce of any highly developed nation like the United States, however, suggests we can no longer assume that completion of secondary and advanced education is for the few. Investment in human resources through education has a greater return rate than capital investment…We cannot return to an economy in which educational opportunities are scarce, but rather must provide enough opportunities that the largest possible proportion of students will acquire the skills and knowledge necessary to sustain the society's growth.[11]

8. Benjamin Bloom, "Mastery Learning," in *Mastery Learning: Theory and Practice*, ed. James H. Block (New York: Holt, Rinehart and Winston, 1971), 49.

9. Ibid., 47–63.

10. Ibid., 50.

11. Ibid., 48.

Bloom's anti-sorting argument came from the belief that the purpose of schooling was to prepare a workforce, and bell-curve sorting was neither efficient nor effective for successful teaching of the greatest number of students. In this example, bell-curve assumptions, even though they are rejected, continue to define the terms of his arguments.

Another fully elaborated critique of sorting appears in Betty Wallace and William Graves's 1995 book, *The Poisoned Apple: The Bell Curve Crisis and How Our Schools Create Mediocrity and Failure*.[12] This book covers a lot of ground, including the usual misunderstandings of statistical representations, several pedagogical disadvantages of bell-curve thinking, the history of the bell curve, and a case study in which an administrator implemented an anti-sorting agenda through policies that included abolishing age-based classroom groupings.[13] Wallace and Graves's thesis is that if teaching is based on bell-curve expectations, then nobody gets a good education. In other words, under a bell curve, teachers are expected to direct their lessons to a fiction called the Average Student, despite the fact that no student actually embodies the characteristics of that statistically generated average. Wallace and Graves argue that teaching based on bell-curve assumptions (1) effectively misses the learning needs of every student, (2) defeats motivation by guaranteeing the expectation of failure, and (3) fosters cheating because of administrative pressures to raise test scores:

> The bell curve serves as both a model and a fitting symbol of an archaic public education system. It describes a broad swath of mediocrity flanked by a sliver of excellence and a ribbon of failure. It is a pattern we've all grown up with, as familiar and seemingly benign as an old schoolhouse bell. But we can no longer afford to be lulled by this warm familiarity. This bell's invitation to learn rings hollow.[14]

Like Bloom in 1971, Wallace and Graves emphasize that bell-curve thinking in education is not an effective basis on which to offer all students maximum learning opportunities. And also like Bloom, Wallace and Graves base their arguments on an assumption that the purpose of schooling is to promote democratic equality, not to offer individuals competitive advantage.[15]

In a third example, Jeannie Oakes, Amy Stuart Wells, Makeba Jones, and Amanda Datnow, longtime outspoken critics of school sorting and tracking, take the position that sorting on the basis of standardized testing is unjust and unreasonable because the required knowledge base and the testing protocols are culturally specific and socially constructed:

12. Betty Wallace and William Graves, *The Poisoned Apple: The Bell-Curve Crisis and How Our Schools Create Mediocrity and Failure* (New York: St. Martin's, 1995).

13. Even though this book engages thoughtfully with ideological, historical, and empirical data about educational success and failure, it is rarely cited in educational research. We suspect one of the reasons for this oversight is that the book is not written in an academic genre. It was cowritten by a practicing teacher and a journalist in a popular, vernacular style.

14. Wallace and Graves, *The Poisoned Apple,* 24.

15. David F. Labaree, "Public Goods, Private Goods: The American Struggle over Educational Goals," *American Educational Research Journal* 34, no. 1 (1997): 39–81.

> the ideology of intelligence is enlisted to make the particular cultural capital (or ways of knowing) of the white and wealthy seem not only more valuable than others, but also the marker of biologically determined ability. This definition of intelligence is reified in the form of standardized tests that measure students' acquisition of this particular cultural capital. This measurement of "ability" provides students from white and wealthy families with considerable advantage, but under the guise of their "natural" abilities, not as a function of their social location.[16]

They cite Howard Gardner's research on multiple intelligences as evidence that intelligence is neither unidimensional nor fixed. Oakes, Wells, Jones, and Datnow target belief in the intelligence ideology as the major obstacle to educational reform:

> The normative and political connections between various conceptions of intelligence and cultural politics emerged strongly in our study, especially as parents and policy-makers articulated their resistance to the schools' reforms. Many educators in the schools we studied struggled mightily to use their own sometimes tenuously altered normative perspectives as wedges to penetrate the fierce political opposition to detracking reforms and the beliefs about intelligence that support rigid track structures.[17]

This argument against sorting is based on assumptions about inequitable distribution of cultural capital in societies; it says that standardized tests and intelligence tests represent the culture of wealthy white people. As such, these tests embed a form of cultural advantage and privilege that unfairly determines and perpetuates a social hierarchy. Again, the assumption behind Oakes et al.'s anti-sorting argument is that the purpose of school should be to equalize the distribution of knowledge and opportunities among people, not to sort people according to some Social Darwinist version of survival of the fittest.

All three anti-sorting arguments described here, although apparently different, are united by their resistance against inequalities in education. They seek to counter inequality through appeals to social justice, social efficiency, and concerns about democratic equality. However, and notably, these three elaborately researched arguments against sorting do not take issue with the existence of bell-curve distribution; rather, they argue that human morality and educational ethics demand that schools make every effort to redress and compensate for inequitable patterns of bell-curve distribution. They argue that if bell-curve distribution is a manifestation of savage competition in nature, then schooling should be about a rational civilizing process in which moral compassion and a sense of justice direct educational practices to overcome natural brutishness.

ANTI-SORTING ARGUMENTS HAVE NOT BEEN EFFECTIVE

The anti-sorting arguments described in the preceding section argue that schools should not sort students because sorting ensures that some people will not get a chance to learn in school. These arguments ring true for people who believe that the purpose of schooling is to promote a more equitable redistribution of resources in society. In educational history, this social-leveling point of view has

16. Jeannie Oakes, Amy Stuart Wells, Makeba Jones, and Amanda Datnow, "Detracking: The Social Construction of Ability, Cultural Politics, and Resistance to Reform," *Teachers College Record* 98, no. 3 (1997): 484.

17. Ibid., 492.

been termed "social melioration," "democratic equality," and "social reconstructionist."[18] The problem is that existing anti-sorting arguments are preaching to the choir.

Not everyone subscribes to a social democratic worldview; some people do not believe that the purpose of schooling is to help redistribute social resources more equitably among all people in a society. Rather, a substantial number of people hold a more capitalistic or individualistic conviction that the purpose of education is to provide individual people with a competitive advantage. In a spirit of Social Darwinism, these people believe in the survival of the fittest, and that the purpose of schooling is to help sort and rank people.

In his sociological study of the conflicting purposes of education, David Labaree calls this point of view the "social mobility" perspective because it holds that education should be a means by which individuals can get ahead of others. Labaree argues compellingly that people in the United States increasingly regard education as a private good and a commodity of personal advantage. Part of the social mobility impulse in education (as a private good) is based in meritocracy and related to testing:

> We can see it in the process of *normalizing judgment*… — rating the performance of each student relative to the others on a bell-shaped curve of student scores — that is embodied in that characteristically American form of testing, the norm-referenced multiple-choice examination. We can see it in the construction of merit-based hierarchies of learners, such as ability groups and curriculum tracks and the whole panoply of other forms of educational stratification. And we can see it in the emphasis on knowledge as private property, which is accumulated for one's personal benefit (under conditions in which sharing is cheating) and which has a value that is measured in the currency of grades, credits, and credentials.[19]

Those who hold that schooling should promote individual advantage are unlikely to support school reform that is designed to provide advantages to the competition. Therefore, people who want schools to serve private goods are likely to oppose educational practices that foster a more equitable distribution of opportunities. This opposition becomes especially strong when it is perceived that such redistribution efforts will cost public money.

Among people who believe that education should provide individuals with competitive advantage, the existing arguments against sorting will not be persuasive because they rest on the assumption that schooling ought to be designed for the public good. We do not hold out much hope that educational capitalists will become socialists in response to the ethical and political appeals of the educational Left. So, as a strategy, the argument of this essay does not have as a prerequisite that people must first be converted from capitalists to social democrats in order to interrogate sorting practices. Rather, the historical approach in this essay is

18. On social melioration, see Kliebard, *The Struggle for the American Curriculum*. On democratic equality, see Labaree, "Essay Review." On social reconstructionism, see Ken Zeichner, "Teachers as Reflective Practitioners and the Democratization of School Reform," in *Currents of Reform in Preservice Teacher Education*, eds. Ken Zeichner, Mary Louise Gomez, and Susan Melnick (New York: Teachers College Press, 1996), 199–214.

19. Labaree, "Public Goods, Private Goods," 57–58.

designed to destabilize the pseudo-scientific basis of sorting practices by showing that bell-curve thinking is not based in natural laws of distribution but rather on historical flukes that became naturalized.

HISTORY OF THE BELL CURVE

> The average man was a fictional being in his creator's eye, but such was his appeal that he underwent a transformation, like Pinocchio or Pygmalion's statue, so that he still lives in headlines of our daily papers.
> — Stephen Stigler, *The History of Statistics*[20]

A bell curve is assumed whenever we are dealing with a probability distribution in samples involving large numbers in a given population. In its applications, the distribution has helped tame hitherto intractable behavior of systems that do not exhibit strict determinacy.[21] In this section, we look at some of those transformations that have extended bell-curve discourse from its original location in mathematics, to statistics, to the domain of scientific thinking involving measurement of human attributes, and eventually to its use as an instrumental tool for rationalizing the measurable outcomes — for example, achievements on standardized tests — of education systems.

The bell curve was invented in the eighteenth century as a way to represent binomial probability. We relate the history of the invention of the bell curve to portray a history of power relations that coalesced into a fundamental assumption about what we can know and how we can study people in a society. Ian Hacking tells the clearest and most succinct story of two different mathematical projects that combined to produce the famous curve attributed to the German mathematician Johann Friedrich Carl Gauss:

> The familiar graphical representation of the idea is the "bell-shaped curve," the Normal distribution of Gaussian distribution that peaks about the mean. There were two routes to this curve. The oldest, written down by De Moivre [friend of Isaac Newton] as early as 1708, obtained it as the limit of a coin-tossing or binomial distribution. We think of a coin being tossed n times, and note the proportion k of heads. After many k-fold trials we obtain a graph showing the number of occasions on which we got 0 heads, 1 head, 2 heads…n heads. The curve will peak around the probability of getting heads with the coin. As n grows without a bound, a Normal distribution results.[22]

Hacking tells us that the second route toward the curve was developed by observational astronomers who used the idea of normal distribution to verify the accuracy of measurements. They measured a distance many times and graphed the results. If most of the measurements clustered around the mean, then the average of those measurements could be considered reliable. Outliers or deviant measurements could be discounted as inaccurate. Obviously, in either the case of the coin toss or the astronomical measurements, the normal distribution describes *probability* based on the construction of a mean.

---

20. Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainly Before 1900* (Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1986), 170. This work will be cited as *HS* in the text for all subsequent references.

21. Ian Hacking, *The Taming of Chance* (Cambridge: Cambridge University Press, 1990).

22. Ibid., 106.

The paraphrasing of the mathematical notion of normal distribution into human sciences was made possible by the specific conditions under which these disciplinary bodies were organized in the nineteenth century. Historical accounts suggest the absorption of the normal curve — with the emergence of people as objects of scientific understanding — as one of the ways in which the emerging human sciences made use of mathematics.[23] We follow this progression beginning with the story of the development of statistics.

The history of statistics compiled by the statistician Karl Pearson is illuminating in this respect: he noted that *Statistik* was nothing more than knowledge of statecraft completely devoid of any trace of mathematics in the times of Gauss, the founder of the idea of normal distribution.[24] *Statistik* had nothing at all to do with numbers. A concomitant development in Great Britain was the founding of the English school of political arithmetic by John Graunt and Sir William Petty. They never used the word *Statistik* nor called their data *statistics*. It was not until a century later that a Scotsman, Sir John Sinclair, adopted the words *statistics* and *Statistik* and applied them to the data and methods of political arithmetic. In mentioning this, Pearson called Sinclair's application of the terms a "bold, barefooted act of robbery."[25] Eventually statistics would be the ground in which the idea of the Average Man took root. In the next step, the normal curve, as booty of this bold and barefooted act of robbery, turned into a positivity ready to be embraced by Adolphe Quetelet in grounding his idea of the Average Man.

Quetelet's Average Man represented a transformation in the understanding of normal distribution from a statement about the regularities of arithmetical probability to an insight into the workings of society.[26] Quetelet inscribed a bell curve within the emerging discourse of social theory that objectified Man as its object. We can see in Quetelet's work an example of the kind of thinking that was characteristic of the mid-nineteenth century and that coalesced to comprise the emerging discipline (or disciplines that fell under the rubric) of social science.[27] According to Hacking:

> Quetelet changed the game. He applied the same curve to biological and social phenomena where the mean is not a real quantity at all, or rather: *he transformed the mean into a real quantity*. It began innocently enough. In a series of works of the early 1830s he gave us "the average man." This did not of itself turn averages — mere arithmetical constructions — into real quantities like the position of a star. But it was a first step.[28]

23. See, for example, Hacking, *The Taming of Chance*; Hanson, *Testing Testing*; and Johann Heilbron, *The Rise of Social Theory*, trans. Sheila Gogol (Minneapolis: University of Minnesota Press, 1995).

24. Karl Pearson, *The History of Statistics in the Seventeenth and Eighteenth Centuries Against the Changing Background of Intellectual, Scientific and Religious Thought: Lectures by Karl Pearson Given at the University College London During the Academic Sessions 1921–1933* (New York: Macmillan, 1978).

25. Ibid., 2.

26. See also Michel Foucault, *The Order of Things: An Archaeology of Human Sciences* (New York: Random House, 1970).

27. See, for example, Heilbron, *The Rise of Social Theory*; and Victor L. Hilts, *Statist and Statistician* (New York: Arno Press, 1981). This rise of social theory was part of a large-scale modernization that also included the invention of the social sciences and the advent of the common school.

28. Hacking, *The Taming of Chance,* 106–107.

After Quetelet, the history of the bell curve is a story of how a strictly mathematical representation of numerical probability got appropriated by social scientists who then began generating numerical descriptions of populations. Data proliferated as more things about people got counted, numerical descriptions of the Average Man were formulated and revised, immigration and industrialization increased in complexity, and statistics increasingly served as a technology by which government offices could rationalize systems of population management, diagnosis, and intervention. The long baking process of history has led us to forget how statistical reasoning was invented, and this amnesia has gradually obliterated the early debates by social scientists about the questionable value of applying mathematical formulae to the understanding of human society.

Yet, we submit that this emerging relation between mathematics and the human sciences was historically fluky and not rationally inevitable. Speaking of this emergence, Michel Foucault asserted that mathematicization was not an inevitable or natural condition of existence of human sciences. In the human sciences, said Foucault, it is possible that the inscription of mathematics was facilitated in a complex way

> by all the modifications that occurred in Western knowledge at the beginning of the nineteenth century. But to imagine that human sciences defined their most radical project and inaugurated their positive history when it was decided to apply the calculation of probabilities to the phenomenon of political opinion…that would be to take a superficial counter-effect for the fundamental event.[29]

The early resistance to bell-curve thinking was very similar to the current resistance to the idea inasmuch as the critique does not challenge the fundamental naturalization of the bell curve. Remarkably, this resistance actually led to the refinement of the techniques for implementing bell-curve thinking.

In the early nineteenth century, the relation of the normal curve to expectations of human behavior was still innovative and debatable as a goal for social science. In fact, in a notable irony of history, one of the strongest arguments *against* using statistics to talk about human society was launched by Auguste Comte, French sociologist and political philosopher, and the founder of positivism. Comte's opposition to statistics was voiced at the same time by Louis Poinsot, the French mathematician, and also on the other side of the English Channel by John Stuart Mill. Poinsot wrote that "the application of this calculus to matters of morality is repugnant to the soul. It amounts, for example, to representing the *truth* of a verdict by a *number*, to thus treat men as if they were dice, each with many faces, some for error, some for truth" (quoted in *HS*, 194, emphasis in original). Mill's description was even more damning when he referred to statistical work as "the real opprobrium of mathematics" (quoted in *HS*, 195). Comte expressed his own disgust with the whole idea of a mathematical basis for social science by calling the enterprise irrational:

> Is it possible to imagine a more radically irrational conception than that which takes for its philosophical base, or for its principal method of extension to the whole of the social sciences,

29. Foucault, *The Order of Things*, 350–351.

> a supposed mathematical theory, where, symbols being taken for ideas (as is usual in purely metaphysical speculation), we strain to subject the necessarily complicated idea of numerical probability to calculation, in a way that amounts to offering our own ignorance as a natural measure of the degree of likelihood of our various opinions? (quoted in *HS*, 194–195)

The application of the normal curve to human sciences was not free of trouble. Careful examinations of Quetelet's Average Man yielded the observation that the whole statistical exercise was circular. That is, in order to count and compare human characteristics, it is first necessary to specify a characteristic in discrete terms, and, second, it is necessary to regard the particular group from whom data were collected as somehow homogeneous — there cannot be too many variables. Consider, for example, the differences that are implied in the following two cases: (1) comparisons across male and female, and (2) comparisons across attached or detached earlobes. In both cases, a continuum (of sexual identities and earlobe attachment) has been converted into a discrete binary. Over time, gender has accrued some discursive homogeneity, whereas earlobe shape generally has not.[30] Quetelet's first studies entailed fairly unambiguously discrete items including births, deaths, heights, and weights. However, it is easy to see that the exercise is circular in that the characteristics themselves are based on preconceived perceptions of differences. In other words, Quetelet's statistical analyses had no means of testing or questioning the existing characteristics. The analysis is bound by its original categories:

> The tool he [Quetelet] had created was too successful to be of use for its original purpose. The fitted distributions gave such deceptively powerful evidence of a stable homogeneity that he could not look beyond them to discover that further subclassification could produce other distributions of the same kind, that some normal distributions are susceptible to dissection into normal components. The method was lacking in discriminatory power; too many data sets yielded evidence of normality. Few new classifications were uncovered; the primary use of the method was as a *device for validating already determined classification schemes*. (*HS*, 215, emphasis added)

Given the strength of the opposition, it was certainly not inevitable that statistical constructions would eventually be transformed into real and essential populational attributes. But the momentum toward the social sciences in the latter part of the nineteenth century was strong enough to overcome the earlier reluctance to describe human characteristics in terms of symbolic entities. The confluence of developments in government, industry, statistics, and the social sciences fostered yet another transformation, namely, from a calculated average to a real thing:

> It was Quetelet's less-noticed next step, of 1844, that counted far more than the average man. He transformed the theory of measuring unknown physical quantities, with a definite probable error, into the theory of measuring ideal or abstract properties of a population. Because these could be subjected to the same formal techniques they became real quantities. This is a crucial step in the taming of chance. It began to turn statistical laws that were merely descriptive of large-scale regularities into laws of nature and society that dealt in underlying truths and causes.[31]

As far as educational theory is concerned, the invention of psychology has served as the vehicle by which statistical analyses became acceptable as tools for

30. In traditional Chinese lore, earlobe length is an indicator of longevity.

31. Hacking, *The Taming of Chance*, 108.

studying human beings.[32] Before the 1880s, researchers regarded various towns and villages as being so idiosyncratic that generalizations across them would be meaningless; there were simply too many extenuating circumstances — from weather, to religion, to custom, language, and local history — for it to be possible to apply classifications across various cases. But conceptual innovations in statistical theory contributed to overcoming these methodological roadblocks. From Quetelet onward, most statisticians accepted this distinction between general and idiosyncratic features, even though "the distinction was blurred and somewhat circular" (*HS*, 256). But the distinction became further blurred as mathematicians developed methodological surrogates for experimental control. Francis Galton's innovation in 1869, which he called "statistics by intercomparison," claimed to be able to measure talent as easily as height, and the argument rested entirely on the logic of analogy.[33]

> Galton turned Quetelet's phenomenon to novel use. If data from the same species arrayed themselves according to this curve and if the unity of the species could be demonstrated by showing that measurable quantities such as stature or examination scores followed the curve, then, once such a species was identified on the basis of measurable quantities, the process could be inverted with respect to qualities that eluded direct measurement!… [T]he use to which the scale was put was clear and somewhat ironic. Where Quetelet had used the occurrence of a normal curve to demonstrate homogeneity, Galton based a dissection of the population upon it. Using this inferred scale, he could distinguish between men's abilities on a numerical scale rather than claim that they were indistinguishable. (*HS*, 271)

Galton's argument by analogy helped make it possible to treat previously immeasurable qualities as discrete entities that could be counted and graphed. Qualities such as cleverness, morality, wit, and civility were appropriated into statistical arrays, and then these qualities gradually became standardized and reified. As they became more common in discourse, it made more sense to regard those qualities as natural. In that way, the qualities of morality and humor were no longer problematic, but rather they became the categorical bases by which the preconceived classifications of race and gender could be validated, and the idea of normal could become naturalized in a particular fashion.

In the history of the bell curve, there was a crucial and profound shift in the meaning of *normal*.[34] The term normal was first used in medical discourse and then transferred to the social sciences following the work of Comte. From the

32. See, for example, Ellen Herman, *The Romance of American Psychology: Political Culture in the Age of Experts* (Berkeley: University of California Press, 1995); and John M. O'Donnell, *The Origins of Behaviorism: American Psychology, 1870–1920* (New York: New York University Press, 1985).

33. The logic of analogies is also related to commensuration, a process that involves using a common metric to compare apparently disparate phenomena. Commensuration is also a basis for bell-curve thinking, with scores on standardized tests providing the common metric for comparison between various apparently disparate categories of human beings. For a thorough discussion on commensuration as an instrument of social thought, see Wendy Nelson Espeland and Mitchell L. Stevens, "Commensuration as a Social Process," *Annual Review of Sociology* 24 (1998): 313–343. For a discussion on commensurability in philosophy of natural sciences, see Thomas S. Kuhn, "Commensurability, Comparability, Communicability," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1982* (Philosophy of Science Association, 1983), 669–688.

34. See Hacking, *The Taming of Chance,* for a thorough discussion of the historical shift in the meaning of "normal."

beginning, *pathological* was defined in opposition to *healthy*. Before 1800, pathology was the central and specified term; anything that was not pathological was assumed to be healthy. The desirable condition, then, was not specified, not circumscribed, and the possibilities for ways of being healthy were theoretically unlimited. After 1800, however, in scientific and social scientific discourse, the term *normal* became the central and specified term. Moreover, the specifications for the term were constituted in a measurable population — either in terms of an average or in terms of a socially defined virtue. In this definition, anything that could not be specifically defined as normal/average was then regarded as pathological/not-normal. In this newer sense, then, the possibilities for being normal are effectively circumscribed as those characteristics that are average or common, and the possibilities for being not normal are limitless. Even the shift in terminology — from *pathological* to *not normal* — connotes the shift in specificity and determination. As Sarah Igo writes,

> Scientific characterizations of "average" or "typical" Americans were a striking phenomenon of the new [twentieth] century. This constituted a shift away from the almost exclusive study of "degenerates, delinquents, and defectives" that had marked nineteenth-century social investigation… [R]igorous inquiry for its own sake into the typicality of everyday practices and opinions was a twentieth-century enterprise.[35]

It may sometimes be supposed that the encroachment of mathematical rationality into the realm of the social is an indication of scientific advance and improvement. However, histories that do not assume this sort of progress have suggested otherwise.[36] We have not argued here that the historical shifts were progressive or regressive, but rather that historical changes were multifaceted, interrelated, and fluky. Nineteenth-century science as constructed in educational discourse did not evolve naturally from previous constructions of science or mathematics. Rather, the constructions of modern science occurred together with other developments of the nineteenth century in bureaucracy, statistics, industry, and organizations.

By the middle of the nineteenth century, it had become usual to think of education as a means to cultivate citizenship; education and citizenship were no longer pulling in opposite directions.[37] The modern political problematic had become how to build a national republic, not by extending influence over territory, but by extending influence over people. One of the tools used to accomplish this was social science, which used laws such as the normal curve to study society and

35. Sarah Igo, *The Averaged American: Surveys, Citizens, and the Making of a Mass Public* (Cambridge, Massachusetts: Harvard University Press, 2007), 11.

36. See, for example, Foucault, *The Order of Things*; Hacking, *The Taming of Chance*; Heilbron, *The Rise of Social Theory*; Peter Wagner, *A Sociology of Modernity: Liberty and Discipline* (New York: Routledge, 1994); and Johann Heilbron, Lars Magnusson, and Bjorn Wittrock, eds., *The Rise of the Social Sciences and the Formation of Modernity: Conceptual Change in Context, 1750–1850* (Boston, Massachusetts: Kluwer Academic, 1998).

37. Compare this stance to that of Jean-Jacques Rousseau, in which being educated was opposed to being a citizen: "If we have to combat either nature or society, we must choose between making a man or making a citizen. We cannot make both. There is an inevitable conflict of aims, from which come two opposing forms of education: the one communal and public, the other individual and domestic." Rousseau, *The Emile of Jean Jacques Rousseau: Selections,* trans. and ed. William Boyd (New York: Teachers College Press, 1956), 13.

determine what could be considered normal. With the development of modern national-scale projects (transportation, communication, commerce, jurisprudence, and arts and sciences), modern educational projects constructed schools as the medium by which potential citizens would be normalized as actual citizens of a republic that had become discursively constructed in the multiple projects of modernity that included bell-curve thinking.

BELL-CURVE THINKING TODAY: STUDENTS ARE AT RISK

> First we make up our claims to truth, then we forget that we made them up, then we forget that we forgot.
> — Friedrich Nietzsche, *On the Genealogy of Morals*[38]

Education is complex, uncertain, and dynamic. However, when the complexities of education are taken up in policy, school administration, and curriculum debates, that uncertain dynamic quality often gets translated — tamed — such that schooling can appear simpler, more predictable, and more static. The social sciences, the bearers of bell-curve thinking, accomplish this translation. The bell curve takes diversity and reduces it to a simple and comprehensible average; it takes a statistical probability and converts it to an expectation that can inform policy; it takes random historical occurrences and imposes patterns of relations on them.

In this section, we examine ways in which the idea of normal in education is shaped by the bell curve. We call attention to three different discursive features to illustrate how widespread and diverse the bell-curve influence has become: (1) the distinction in educational assessments between norm-referenced and criterion-referenced reporting of test results, (2) the effects of polls and surveys in educational referenda and research, and (3) the growing use of averages and normal distributions not just to represent human behavior but to establish ideal norms for how people should behave. These three examples illustrate how the bell curve is functioning in present circumstances. We conclude by drawing on Robert Castel's distinction between dangerousness and risk to suggest ways in which bell-curve averages get projected into the future as expectations for how people should act.

NORM-REFERENCED AND CRITERION-REFERENCED TEST-SCORE REPORTS

We can see bell-curve assumptions about normalcy in the definitions that distinguish criterion-referenced test scores from norm-referenced test scores. In common understandings, a criterion-referenced score reflects the test taker's performance against a degree of mastery in a selected domain. In contrast, a norm-referenced score reflects the test taker's performance against the performances of other test takers. This distinction may be taken to imply that criterion-referenced tests are not based on social norms. However, the ability to distinguish between norm-referenced and criterion-referenced scores requires an element of forgetting.

38. Friedrich Nietzsche, *On the Genealogy of Morals*, trans. Douglas Smith (1887; repr. London: Oxford University Press, 1996).

In this section we suggest three perspectives that help show the ways in which criterion-referenced tests are also norm-referenced tests.

First, individual test items are developed according to social norms. For any criterion-referenced test, the particular test questions were derived from sociocultural assumptions about what is important to know. All test items are written and chosen through a process of test development in which results from pilot tests get compared to the results of previously established tests, and new test items are continually modified until the new test results correlate with the old test results. In that way, the scoring of criterion-referenced tests is reported on the basis of a percent of right and wrong answers, but those answers are themselves based on social norms. From a historical point of view, criterion-referenced tests are also norm-referenced tests because the criteria are all socially constructed.

Second, there are really no criterion-referenced tests, but only a criterion-referenced manner of interpreting particular raw scores. If the scores are seen in relation to a norm, the test assumes a norm-referenced tenor. In contrast, if scores are seen in relation to a criterion, the background shifts, and the scoring appears to be criterion referenced. That is, both criterion as well as norm could be referenced within the same test.

To understand the ways in which this conflation works, consider an example from gym class in which a fitness test regards the number of sit-ups performed by students as a measure, or *construct*, of abdominal muscle strength. There are charts, tables, and graphs that tell us how many sit-ups a child in the fourth grade should be able to do. Therefore, a particular scientific discourse on abdominal muscles tells us how strong children should be at various ages. The test is administered in schools, and all the students are made to do sit-ups — the measure of strength — in every grade from the third to the eighth. The numbers we get from this test are simply raw scores. The norm referencing for such a test will require us to set a norm arbitrarily. So if we assume it to be set at the *mean score* of the eighth grade students, then *normality* will be defined with reference to that mean score. A particular student in eighth grade who scores in the bottom percentile may have the same strength as the mean for fourth grade students. In that case, the student is scoring like an average fourth grader, or much weaker than the normal. But the raw score that student obtained on the same measure could also be seen in terms of a criterion. So if the criterion for passing was set at ten sit-ups for eighth graders and he did nine, he failed. What is important in this example is that the same score was processed through a criterion and a norm.

Third, the intelligence quotient is an example of another kind of conflation of norm-referenced and criterion-referenced results. By definition, an IQ score of 100 is calibrated to represent the fiftieth percentile of test takers and thus is norm-referenced; so for an IQ measurement, there is no distinction between a criterion- and norm-referenced score. In these three examples we suggest that a clear distinction between criterion- and norm-referenced tests is an example of forgetting, and this forgetting contributes to the ways the bell curve has shaped the idea of normal.

SURVEYS AND POLLS

Another discursive feature shaping bell-curve assumptions is the administration of, participation in, and reporting of surveys and opinion polls.[39] Surveys are generally regarded to be reasonable instruments by which we can quantify and identify knowledge about human thought and behavior. Of course, much polling done for purposes of political campaigning and commercial advertising is intentionally skewed and unscientific, and we will not consider those trivial examples here. Rather, we are referring to research surveys and polls that are constructed with meticulous care for objectivity, validity, and reliability. We argue that polls and surveys recursively shape what is perceived to be normal or average. Ironically, perhaps, the more sophisticated the survey, the more profound its potential impact on the construction of normal.

Igo's recent book offers a fully documented historical analysis of the ways in which surveys have shaped assumptions about normality in the twentieth-century United States. She traces the inception and development of survey practices by focusing on the histories of the Roper, Gallup, and Kinsey polls. Her analysis shows how survey results had effects not only of assimilation and homogenization, but the publication of social averages also served to bolster antiestablishment movements:

> Survey technologies never worked simply to normalize their subjects in the service of a consensus society, however. This would not account for their widespread appeal. Surveys could also encourage and give new weight, through numbers, to nonnormative habits, beliefs, and identities. Social statistics, highlighting both inclusion and exclusion, prompted some to imagine themselves into new collectives or to forge a minority consciousness.[40]

The interesting emphasis in Igo's contribution is that bell-curve definitions of normal do not necessarily promote assimilation and homogeneity. Rather, the discursive impact of bell-curve thinking can be mobilized with different kinds of values with respect to what is normal.

FROM REPRESENTATION TO EXPECTATION

The history of the bell curve tells us that it began as a mathematical design for representing binomial data. Then the bell curve transformed into a way to represent nondiscrete qualities (like morality and intelligence). In this section, we focus on the most recent transformation of the bell curve, namely, a shift from a means of representing to a device for establishing expectations for how people should behave.

Claude Steele and Joshua Aronson's famous study that has become known as the "stereotype threat" serves as one example of this process of establishing expectations.[41] In their controlled-experiment study, standardized tests were

39. For an explanation of the connection between bell-curve thinking and polling practices, see "How Polls Work," available at http://www.notrain-nogain.org/Train/Res/Num/how.asp.

40. Igo, *Averaged American*, 285.

41. Claude M. Steele and Joshua Aronson, "Stereotype Threat and the Intellectual Test Performance of African-Americans," *Journal of Personality and Social Psychology* 69, no. 5 (1995): 797–811.

administered to different groups of students. Just before students began the test, the treatment groups were told that the test results were going to be used in a study about demographic differences. The control groups were told that the test results were going to be used in a study about mental processing. In the treatment groups, when the focus was on race and gender, the test scores for African American and Latino students were significantly lower than in the control cases. In Steele and Aronson's study, the test scores of high-achieving African American students were lowered in conditions when it was merely suggested that negative racial stereotypes could affect the test they were about to take.

Another way to think about the relation between stereotypes and behavior has been what Hacking calls the "looping effect" of human kinds.[42] Hacking introduces his concept of looping in the context of a book about causal reasoning. By looping, Hacking means that as categories get constructed from averages, causal connections are made to help "explain" the categories. In discourse, these categories and causes about human kinds become part of knowledge, of which we humans are both subjects and objects. Hacking provides examples from psychiatry and pediatrics to illustrate that the relation between classification and the attribution of cause is a chicken-and-egg loop:

> Which comes first, the classification or the causal connections between kinds?… [T]o acquire and use a name for any kind is, among other things, to be willing to make generalizations and form expectations about things of that kind… The kind and the knowledge grow together… [In the case of pediatric X-rays and child abuse] cause, classification, and intervention were of a piece.[43]

Similarly, in the case of Steele and Aronson's stereotype threat, the test scores of the African American students can be seen as a reiteration of stereotypes in a kind of self-fulfilling prophecy.

Hacking's historical treatment of looping introduces another level — a discursive level — of reiteration to processes of bell-curve thinking. That is, not only are categories reiterative in an analytical sense (the reasoning is circular, as we just explained), but categories are also reiterative in a discursive/historical sense (repetition of labels has an effect on possibilities for identity over time). Assumptions about what is normal are shaped by habit and custom, and now the bell-curve definition of normal has itself become a habitual expectation that continues to validate belief in itself.

Implications and Determinations

In this last section we focus on the normalization that is made possible when schooling is seen as an example of modern social management and administration. We understand why bell-curve thinking has been attractive for administration and management in educational endeavors. Bell-curve generated averages and classifications help to clarify and simplify a confusing morass of dynamic diversity for

42. Ian Hacking, "The Looping Effects of Human Kinds," in *Causal Cognition: A Multidisciplinary Debate*, ed., Dan Sperber, David Premack, and Ann James Premack (Oxford: Clarendon Press, 1995), 351–394.

43. Hacking, *The Taming of Chance*, 361.

purposes of policy making. Bell-curve thinking can help provide a rationale for routine administrative practices such as putting students into classes (whether or not the classes are homogeneously or heterogeneously sorted), levying and spending tax dollars for particular kinds of school reforms, and providing curricula that efficiently prepare students for life beyond school.

We argue here that the bell curve has also shaped the idea of normal in education by functioning as a technology of risk management. The successful management of educational endeavors (including setting policies, building schools, setting standards, and the like) rests on the ability to predict what is going to happen in the future in order to anticipate problems and take proper precautions. This sort of thinking constitutes rational planning for purposes of managing schools at all levels. The bell curve contributes to rational planning by creating the Average Student. If schooling practices can be seen to serve the needs of the Average Student, then those practices can be regarded as reasonable by the general public.

However, as a device for thinking about educational enterprises, the Average Student gets transformed from an arithmetic mean to a real entity — just as Quetelet's quantities did in the 1830s. In discussions about educational policies, the Average Student is not treated as an arithmetic construct — a fiction. When the Average Student becomes the Normal Student, no real student can be normal. Under bell-curve standards of normalcy, all students are deviant. All students are at risk and therefore in need of normalization.

We now consider what happens when we take the Average Student and use that fiction as a basis for establishing educational expectations. Recalling Hacking's looping effect, we now connect that looping with a notion of risk management.

An elegant explication of the distinction between dangerousness and risk is provided by Robert Castel.[44] According to him, dangerousness happens in face-to-face relations; danger is a historically specific and (relatively) unmediated relation. Risk, on the other hand, is a future projection. It is a statement of probability and the product of a rational calculation. Risk is mediated because the projected judgment about any given circumstance is made on the basis of bell-curve thinking, specifically, statistical averages, abstract probabilities, and generalized principles. In describing the historical shift in governance from dangerousness to risk, Castel writes:

> The essential component of intervention no longer takes the form of the direct face-to-face relationship between the carer and the cared, the helper and the helped, the professional and the client. It comes instead to reside in the establishing of *flows of population* based on the collation of a range of abstract factors deemed liable to produce risk in general.[45]

Castel focuses on the historical shifts in psychiatry and social work, but a parallel can be drawn with education, especially because education and psychiatry

44. Robert Castel, "From Dangerousness to Risk," in *The Foucault Effect: Studies in Governmentality*, ed. Peter Miller (Chicago: University of Chicago Press, 1991), 281–298.

45. Ibid., 281 (emphasis in original).

were closely tied at the beginning of the twentieth century through the work of eugenicists.[46] Education, psychology, and social work were facets of the emerging social sciences. The approach to education — and the social sciences in general — became conceived in terms of "preventive prophylaxis."[47] Modernization of education was a process by which personal subjective evaluations were replaced with abstract objective evaluations in which the idea of normal became the standard. Pupils began to be thought of in terms of classes, citizens became inscribed with the attributes of populations, teaching was classified according to methods, behaviors became generalized in terms of psychological propensities, and schools were organized into an educational system.

The point is that statistical averages — including those generated by bell-curve representations — became modern, rational, and reliable standards. Standards gradually became formalized in all sorts of educational discourse, from policy to curricula to assessments. This formalization contributed to the ease with which it became possible to understand the purpose of education to be the normalization of people. We can take Castel's description of psychiatry as analogous to education:

> The site of diagnostic synthesis is no longer that of the concrete relationship with a sick person [student], but a relationship constituted among the different expert assessments with make up the patient's [student's] dossier. Already here there is the shift from presence to memory, from the gaze to the objective accumulation of facts. The resulting situation…might be called a transition from a clinic [school] of the subject to an "epidemiological" clinic [school], a system of multifarious but exactly localized expertise which supplants the old doctor-patient [teacher-student] relation. This certainly does not mean the end of the doctor [teacher], but it does definitely mark a profound transformation in medical practice [schooling practices].[48]

Applying this analysis to education, we can see that the history of schooling entailed a process of regulation and governance that presumed to take control of the future by means of preventive maintenance. The bell curve is one of the technologies by which that governance has been understood and managed. At the same time, that "system of multifarious but exactly localized expertise" has increased in magnitude and density through the accumulation of repetitions at capillary

46. Relevant to education, developmental psychology has its roots in the eugenics movement, although those ties have — understandably — been eliminated from many histories of developmental psychology. See, for example, Ira Bretherton, "The Origins of Attachment Theory: John Bowlby and Mary Ainsworth," in *A Century of Developmental Psychology*, ed. Ross D. Parke, Peter A. Ornstein, John J. Rieser, and Carolyn Zahn-Waxler (Washington, D.C.: American Psychological Association, 1994), 431–471; David Elkind, "David Elkind," in *A History of Developmental Psychology in Autobiography*, ed. John D. Hogan (Boulder, Colorado: Westview Press, 1996), 71–83. For other kinds of histories of psychology, see, for example, Kurt Danziger, *Constructing the Subject: Historical Origins of Psychological Research* (Cambridge: Cambridge University Press, 1990); Herman, *The Romance of American Psychology*; Stephen J. Gould, *The Mismeasure of Man*, rev. ed. (New York: W.W. Norton, 1996); O'Donnell, *The Origins of Behaviorism*; and Valerie Walkerdine, "Developmental Psychology and the Child-Centered Pedagogy: The Insertion of Piaget into Early Education," in *Changing the Subject: Psychology, Social Regulation, and Subjectivity*, ed. Julian Henriques, Wendy Holloway, Cathy Urwin, Couze Venn, and Valerie Walkerdine (New York: Methuen, 1984), 154–210.

47. Castel, "From Dangerousness to Risk," 285.

48. Ibid., 282.

levels of thought and behavior in several social sectors, including education.[49] We can see this accumulation in the increase in frequency and diversity of tests (for example, aptitude, achievement, proficiency, placement, personality, learning style, self-esteem, and teacher-perceiver) that are administered in schools. In a sense, then, as rationalization and essentialization became unquestioningly assumed, the idea of normal has become formalized and normalized.[50]

The issue is not about whether the bell curve is good or bad — the terms of the debate reiterated by both critics and advocates of bell-curve thinking — but it is about how bell-curve thinking acquires its historical truth function when repeated in the practices of averaging, in norm and criterion referencing, in polls and surveys, and so on. The bell curve functions to tame chance by projecting order on chaos, manageability on the otherwise unmanageable and dangerous, and policy on what would otherwise remain un-policed. Every time bell-curve thinking is repeated (in test results, in statistical analyses of test items, in predictions of "normal" distributions, by educational record keeping, and by people's self-identifications on bureaucratic forms), the concocted notion of normal is reinscribed and reinforced. The bell curve is deployed ostensibly as a means to prevent, manage, or at least keep track of failure. However, by locking in the idea of normal-as-average, bell-curve thinking guarantees some degree of failure in all educational projects and makes sorting appear normal.

49. See also Hacking, *The Taming of Chance*; Thomas S. Popkewitz, *A Political Sociology of Educational Reform: Power/Knowledge in Teaching, Teacher Education, and Research* (New York: Teachers College Press, 1991); and Nikolas Rose, *Governing the Soul: The Shaping of the Private Self* (London: Routledge, 1990).

50. In Foucault's terms, discursive formation can be said to have crossed the threshold of formalization. Michel Foucault, *Archaeology of Knowledge* (New York: Pantheon, 1972).