DEVANE LECTURE

Would you please join me in a brief fantasy? The year is 2020. Richard Levin has just retired after a long and brilliant career as President and has declared this "2020 The Year of Perfect Vision." Every last building is rebuilt and shining, the students are even more precocious, accomplished, and unionized than they are today (2001), US News and World Report and Consumer Report (now merged) have ranked Yale University Number 1 across the board---up there with the very best hotels, luxury automobiles and lawnmowers. Well,…. nearly across the board. It seems that the quality of the faculty, as reflected in the all-important rankings, has slipped. Yale's competitors are shaking their heads at the decline. Those who know how to read between the lines of apparently serene "Yale Corporation" pronouncements, can detect a rising, but, of course, still decorous, panic.

One sign of concern can be read from the selection of President Levin's successor, Condoleeza Rice, the retired Secretary of State who most recently led a no-nonsense, business-like, streamlining of the Ford Foundation. Yes, she is the first woman of color to lead Yale. Of course, four other Ivy League schools are also headed by women of color. Not surprising, inasmuch as Yale has always followed the New England farmer's rule: "Never be the first person to try something new, nor the last."

On the other hand, President Rice wasn't chosen for the symbolism; she was chosen for the promise she represented: the promise of leading a thorough-going restructuring of the faculty using the most advanced quality-management techniques: techniques perfected from the crude beginnings the Grandes Ecoles of Paris in the late 19th century, embodied in Robert MacNamara's revolution in the Department of Defense in the 1960s, Margaret Thatcher's managerial revolution in British social policy and higher education in the 1980s, further developed by the World Bank, and brought to near perfection, so far as higher education is concerned, by the Big Ten Universities and making its way, belatedly, to the Ivy League.

We know, from confidential sources among the members of the Yale Corporation, how she captivated them in her 'job interview'. She said that she admired the judicious mix of feudalism – in politics--and capitalism ---in financial management--- that Yale had managed to preserve. It suited perfectly the reforms she had devised---as did Yale's long tradition of what has come to be celebrated as "participatory autocracy" in faculty governance.

But it was her comprehensive plan for massively improving the quality of the faculty—or, more accurately, to improve its standing in the national rankings, that convinced the Corporation that she was the answer to their prayers.

She excoriated Yale's antiquated practices of hiring, promoting and tenuring faculty. They were, she said, subjective, medieval, unsystematic, capricious and arbitrary. These customs, jealously guarded by the ageing—largely white male-- Mandarins of the faculty, whose average age now hovered around 80, were, she claimed, responsible for Yale's loss of ground to the competition. They produced, on the one hand, a driven, insecure, junior faculty who had no way of knowing what the criteria of success and promotion were beyond the tastes and prejudices of the seniors in their department and, on the other hand, a self-satisfied, unproductive, oligarchy of gerontocrats heedless of the long run interests of the institution.

Her plan, our sources tell us, was beguilingly simple. She proposed using the scientific techniques of quality evaluation employed elsewhere in the academy, but to implement them, for the first time, in a truly comprehensive and transparent fashion. The scheme hinged on the Citation Indices: The Arts and Humanities Citation Index, the Social Science Citation Index, and the granddaddy of them all, the Science Citation Index. Sure, these counts of how often one's work was cited by others in the field were consulted from time to time in promotion reviews, but as President Rice, she proposed to make this form of objective evaluation systematic and comprehensive. The citation indices, she stressed, like the machine counting of votes, play no favorites; they are incapable of conscious or unconscious bias; they represent the only

impersonal metric for judgments of academic distinction. They would henceforth be the sole criterion for promotion and tenure. If she succeeded in breaking tenure, it would also serve as a basis for automatically dismissing tenured faculty whose sloth and dimness prevented them from achieving annual citation norms. (ACN for short).

In keeping with the neo-liberal emphasis on transparency, full-public disclosure, and objectivity, President Rice proposes a modern, high-tech, academic version of Robert Owens's factory scheme at New Lanark. The entire faculty is to be outfitted with digitalized beanies. As soon as they are designed—in Yale's distinctive blue-and-white, and can be manufactured under humane, non-sweatshop conditions, all faculty will be required to wear them on campus. The front of the beanie, across the forehead, will consist in a digital screen, rather like a taximeter, on which will be displayed the total citation count of that scholar in real time. As the fully automated citation recording centers register new citations, these citations, conveyed by satellite, will be posted automatically to the digital read-out on the beanie. Think of a miniature version of the constantly up-dated world population count once available in lights in Times Square. Let's call it, Public Record of Digitally Underwritten Citation Totals (PRODUCT, for short). Rice conjured a vision of the thrill students would experience as they listened, rapt, to the lecture of a brilliant and renowned professor whose beanie, while she lectured, was constantly humming, the total citations piling up before their very eyes. Meanwhile, in a nearby classroom, students worry as they contemplate the blank readout on the beanie of the embarrassed professor before them. How will their transcript look when the cumulative citation total of all the professors from whom they have taken courses is compared with the cumulative total of their competitors for graduate or professional school? Have they studied with the best and brightest?

Students will no longer have to rely on the fallible hearsay evidence of their friends or the prejudices of a course critique. The numerical "quality grade" of their instructor will be there for all to see, and to judge. Junior faculty will no longer need to fear the caprice of their senior colleagues. A single, indisputable standard of achievement will, like a batting average, provide a measure of quality and an unambiguous target for ambition. For President Rice, the system solves the perennial problem of how to reform departments that languish in the backwaters of their disciplines and become bastions of narrow patronage. This publicly accountable, transparent, impersonal measure of professional standing shall henceforth be used, in place of promotion and hiring committees.

Think of the clarity! A blue-ribbon panel of distinguished faculty (chosen by the new criterion) will simply establish several citation plateaux: one for renewal, one for promotion to term associate, one for tenure, and one for post-tenure performance. After that, the process will be entirely automated once the beanie technology is perfected. Imagine a much-quoted, pace-setting Political Science Professor, Harvey Writealot, lecturing to a packed hall in WLH. Suddenly, because an obscure scholar in Arizona has just quoted his last article in the Journal of Recent Recondite Research (RRR) and, by chance, that very quote is the one that puts him over the top, the beanie instantly responds by flashing the good news in blue-and-white and playing Boola-Boola. The students, realizing what has happened, rise to applaud their professor's elevation. He bows modestly, pleased and embarrassed by the fuss, and continues the lecture—but now with tenure. The console on the desk of President Rice's office in Woodbridge Hall, tells her that Harvey has 'made it' into the magic circle on his own merits and she, in turn, sends him a message of congratulations broadcast through the beanie by text and voice. A new, distinctive, "tenure-beanie" and certificate will follow shortly.

Members of the Corporation, understanding instantly how much time and disputation this automated system could save and how it could catapult Yale back into the faculty-ratings chase, set about refining and perfecting the technique. One suggested having a time-elapse system of citation depreciation; each year's citations will lose one eighth of their value with each passing year. An eight-year-old citation evaporates in keeping with the pace of field development. Reluctantly, one member of the Corporation suggested that, for consistency, there be a minimal plateau for retention, even of previously tenured faculty. She acknowledged that the image of a bent professor's citation total degrading to the dismissal level in the middle of a seminar is a sad spectacle to contemplate, another suggested that the beanie, in this case, could simply be

programmed to go completely blank, though one imagines the professor could read his fate in the averted gaze of his students.

Thank you for indulging my fantasy. I think you see where I'm headed. I'm abandoning the fantasy but continuing with a polemic. The point I want to make is reasonably straightforward. It is that democracies, particularly mass democracies like the United States that have embraced meritocratic criteria for elite selection and mass opportunity, are tempted develop impersonal, objective, mechanical, measures of quality. Why? The short answer is that there are few social decisions as momentous for individuals and families as the distribution of life-chances through education and employment or as momentous for communities and regions than the distribution of public funds for vast public-works projects. The seductiveness of such measures is that they all turn measures of quality into measures of quantity, thereby allowing comparison across cases with an apparently single metric. Whether they take the form of the Social Science Citation Index, the Research Assessment Exercise (RAE, in Britain), the Scholastic Aptitude Test (SAT), Cost-Benefit Analysis CBA), they all follow the same logic.

I hope to sketch that logic and argue that it is irremediably and fatally flawed. The first and most obvious problem with such measures is that they are often invalid: that is, they rarely measure the quality we believe to be at stake with any accuracy. This issue, on which most ink has been spilled, will not detain us long.

The second and greatest flaw of these administrative techniques is that they function as a vast "anti-politics machine", sweeping vast realms of legitimate public debate out of the public sphere and into the arms of technical, administrative, committees. Far from eliminating politics, such techniques merely bury a vital politics in a series of conventions, measures, and assumptions that escape public scrutiny and public education. They stand in the way of potentially bracing and instructive debates about social policy, the meaning of intelligence, the selection of elites, the value of equity and diversity, and the purpose of economic growth and development. They are, in short, the means by which technical and administrative elites attempt to convince a skeptical public---while excluding them from the debate--- that they play no favorites, take no obscure discretionary action, and have no biases but are merely making transparent technical calculations. At a larger level, I shall argue that they are the hallmark of a neo-liberal political order in which the techniques of neo-classical economics have, in the name, of scientific calculation and objectivity, come to replace, other forms of reasoning. Whenever you hear someone say "I'm deeply invested in him/her" or refer to social or human "capital" or, so help me, refer to the "opportunity cost" of a human relationship, you'll know what I'm talking about.

The third fatal flaw is that even if the measure, when it was first devised, were a valid measure, its very existence typically sets in motion a train of events that undermine its validity. Let's call this a process by which "a measure colonizes behavior", thereby negating whatever validity it once had. An historical example will clarify what I mean. The officials of the French absolutist kings sought to tax their subjects' houses according to size. They seized on the brilliant device of counting the windows and doors of a dwelling. At the beginning of the exercise, the number of windows and doors was a nearly perfect proxy for the size of a house. Over the next two centuries, however, the "window and door tax," as it was called, impelled people to reconstruct and rebuild houses so as to minimize the number of apertures and thereby reduce the tax. What started out as a valid measure became an invalid measure.

If time permits, I hope to extend my argument to the policy sciences, and particularly the technique of Cost-Benefit-Analysis by which the vast majority of projects and development schemes in the West and in Asia, Africa, and Latin America are justified. I will begin, however, closer to home, with education and, particularly, with the Social Science Citation Index, the Research Assessment Exercise of Great Britain, and the Scholastic Aptitude Test.

The Science Citation Index (SCI), founded in 1963 and granddaddy of all citation indices, was the brainchild of Eugene Garfield. Its purpose was to gauge, to measure the scientific impact of, say, a particular research paper and by extension, a particular scholar or research laboratory, by the frequency with which a paper was, in turn, cited by other research scientists. Why not? It sure beat relying on informal reputations, grants, the obscure embedded hierarchies of established

institutions, let alone the sheer productivity of a scholar. More than half of all scientific publications, after all, seem to sink without a trace; they aren't cited at all, not even once! Eighty percent are only cited once, ever. The Science Citation Index seemed to offer a neutral, accurate, transparent, disinterested, and objective measure of a scholar's impact on subsequent scholarship. A blow for merit. And so it was, at least initially, compared to the structures of privilege and position it claimed to replace.

It was a great success, not least because it was heavily promoted; let's not forget that this is a for-profit business! Soon it was pervasive: used in the award of tenure, to promote journals, to rank scholars and institutions, in technological analyses and government studies. Soon the Social Science Citation Index followed and, after that, could the Arts and Humanities Citation Index be far behind?

But was it valid? What precisely did it measure? The first thing to notice is the computer-like mindlessness and abstraction of the data gathering. Self-citations counted, adding auto-eroticism to the normal narcissism that prevails in the academy. Negative citations, "X's article is the worst piece of research I have ever encountered". Score one for X! As Mae West said, "There's no such thing as bad publicity; just spell my name right." Citations found in books, as opposed to articles, are not canvassed. More seriously, what if absolutely NO ONE EVER READ the articles in which a work was cited, as was often the case? Then there is the provincialism of the exercise; this is, after all a massively English language, and hence Anglo-American operation [as is, of course, this series of lectures]. Garfield claimed that "French science is provincial because of its reluctance to accept English as THE scientific language". In the social sciences, this is preposterous on its face, but it is true that the translation and sale of your work to a hundred thousand Chinese, Brazilians, or Indonesians will add nothing to your Social Science Citation Index standing unless they record their gratitude in an English language journal or one of the handful of foreign language journals included in the magic circle.

Notice, too that the index must, as a statistical matter, favor the specialties that are the most heavily-trafficked, that is to say mainstream research or, in Kuhn's terms. "normal science". Notice finally that the "objectified subjectivity" of the Social Science Citation Index also is supremely presentist. What if a current line of inquiry is dropped as a sterile line of inquiry three years hence? Today's wave, and the index blip it creates, may still have allowed our lucky researcher to surf to a safe harbor despite her mistake. There is no need to belabor these shortcomings of the SSCI further. They serve only to show the inevitable gap between measures of this kind and the underlying quality they purport to assess. The fact is that many of these shortcomings could, at least in principle, be rectified by reforms and elaborations in the procedures by which the index is constructed. In practice, however, the more schematically abstract and computationally simple measure is preferred for its ease of use and, in this case, lower cost.

The anti-politics of the Social Science Citation Index consists in substituting a pseudo-scientific calculation for a healthy debate about quality. The real politics of a discipline- its worthy politics anyway- is precisely the dialogue about standards of value and knowledge. I entertain few illusions about the typical quality of the dialogue. Are there interests and power relations at play? You bet. They're ubiquitous. There is, however, no substitute for this necessarily qualitative and always-inconclusive discussion. It is the life-blood of a discipline's character, fought out in reviews, classrooms, roundtables, debates, and decisions about curriculum, hiring, and promotion. Any attempt to curtail that discussion by, for example, Balkanization into quasi-autonomous sub-fields, rigid quantitative standards, or elaborate score-cards tends simply to freeze a given orthodoxy or division-of-spoils in place.

My fun at the expense of the SSCI may seem a cheap shot. The argument I'm making, however, applies to ANY quantitative standard, rigidly applied. Take the apparently reasonable "two-book" standard often applied here in tenure decisions. How many scholars do you know whose single book or article has generated more intellectual energy than the collected works of other, quantitatively more "productive", scholars? The

commensurating device known as the “tape-measure” may tell us that a Vermeer portrait and a cow plop are both ten inches across; there, however, the similarity ends.

Once in place, any Index used to distribute rewards or punishments begins to circle back and colonize behavior. Thus, I have been told that there are “rings” of scholars who have agreed to cite one another routinely and thereby raise their citation rating. Outright conspiracy of this kind is but the most egregious version of a more important phenomenon. Simply knowing that, as in the “Window and Door Tax”, that the Citation Index COUNTS, that is, can make or break a career, then exerts a subtle or not-so-subtle influence on professional conduct: for example, the advantages of mainstream methodologies and populous sub-fields, the choice of journals, the incantation of a field’s most notable figures are all encouraged the incentives thereby conjured up. This is not crass Machiavellian behavior; I’m pointing instead to the constant pressure at the margin to act “prudently”. The result, in the long run, is a selection pressure, in the Darwinian sense, favoring the survival of those who meet or exceed their audit quotas.

A citation index is not merely an observation; it is a force in the world, capable of generating its own observations. Social theorists have been so struck by this “colonization” that they have attempted to give it a law-like formulation. Goodhart’s version is: “Any observed regularity will tend to collapse, once pressure is placed on it for control purposes.” Harkin reformulates it more succinctly: “Every measure which becomes a target becomes a bad measure.” Matthew Light my research assistant and collaborator on this topic, puts it very lucidly: “An authority sets some quantitative standard to measure a particular achievement; those responsible for meeting that standard, do so, but not in the way which was intended.”

Here’s an innocuous example. Customer-relations personnel at Amazon.com were rated, on a monthly basis, on their average “call-resolution-time”: i.e. on how long it took to serve the customer or answer their query or complaint. One employee, noting late in the month that his average was well below the standard, quickly improved his average by answering call after call with, “Hello, Thank you for calling Amazon.com” and then immediately hanging up.

Here’s a not-so-innocuous example. The desire for measures of performance that were quantitative, impersonal, and objective were, of course, integral to the management techniques brought from General Motors to the Pentagon by “whiz-kid” Robert McNamara and applied to the war in Indochina. In a war without clearly demarcated battle-fronts, how could one gauge progress? McNamara told General Westmoreland, “General, show me a graph that will tell me whether we are winning or losing in Vietnam.” The result was at least two graphs: one, the most notorious, was an index of attrition, in which the “body-counts” of confirmed enemy personnel killed-in-action were aggregated. Under enormous pressure to show progress and knowing that the figures influenced promotions, decorations, and rest-and-recreation decisions, the body counts swelled. Any ambiguity between civilian and military casualties was elided; virtually all dead bodies became enemy military personnel. Soon, the total of enemy dead exceeded the known combined strength of the so-called Viet Cong and the North Vietnamese forces troop levels. Yet, in the field, the enemy was anything but exhausted.

The second index was an effort to take the measure of civilian sympathies in the campaign to Win Hearts and Minds (WHAM). The Hamlet Evaluation System was at its core—every one of southern Vietnam’s 12,000 hamlets was classified, according to an elaborate scheme: “pacified”, “contested”. or “hostile”. Pressure to show progress was again unrelenting. Ways were found: by fudging figures, by creating ‘paper’ self-defense militias that would have make Tsar Catherine’s Minister Potempkin proud, by statistically ignoring incidents of insurgent activity, in order to have the graph show improvement. Outright fraud, though not rare, was less common than the understandable tendency to resolve all ambiguities in the direction the incentives for evaluation and promotion led. Gradually, it seemed, the countryside was being pacified.

McNamara had created an infernal system that not only produced a mere simulacrum, a "command performance" as it were—of legible progress, but also blocked a wider-ranging dialogue about what might, under these circumstances, represent progress. They might have heeded a real scientist's words: Einstein's, "Not everything that counts can be counted and not everything that can be counted, counts."

AUDITS AND COMMENSURATION

Originally, and literally, "a hearing", the audit was an oral account by a vassal or serf to the lord or his agent of the services and dues rendered. It came, later, to mean the checking of "accounts" – especially of quantifiable activity- against other documentary evidence (receipts, bills, tax statements, contracts) for accuracy. At the limit, it implied tracing every transaction back to its origin for verification. Audits, then, were from the beginning born of a fear of misrepresentation and fraud. As with other forms of accountability and quantification, they were, as Theodore Porter so aptly puts it, "technologies designed to overcome distrust and distance". They serve the rhetorical and political functions of convincing a possibly skeptical outsider that the affairs of the party or institution is fairly represented here; it is a guarantee that supplements or, in some cases, substitutes for trust founded upon reputation, kinship, religious ties, and friendship. Double-entry bookkeeping, a Renaissance innovation was designed as much to improve the social standing of merchants- demonstrating their probity and professionalism—as actually to keep accurate records. The books of the British East India Company were kept only sporadically. Auditing, as standard practice of enterprises and public bodies, came into its own with the modern corporation's separation of ownership and management and the rise of mass democracy. Distrust and distance are at work here again.

Against this crude background, how might we explain the huge, one might say hegemonic, influence of 'quantitative representations of QUALITIES, such as the Scholastic Aptitude Test (SAT) and Cost-Benefit Analysis? The conjunction of high-modernist faith in the efficacy of scientific and technical management on the one hand and, again, mass democracy, on the other, seem central to their success.

High modernism provides an ideological climate in which numerical assessments of quality and decision-making techniques: such as the SAT and Cost-Benefit-Analysis respectively, can flourish. As a muscle-bound version of modernism, technical [not literary or artistic] high modernists believe that the great advances in science and technology provided unitary and, for the moment, definitive answers to the design of society, to social policy, to how a city ought to look and function, to how infants ought to be bathed, to how social rewards ought to be distributed, and to how crops ought to be grown. Nothing escaped the hubris of the high-modernist lens: neither the re-designing of homo-sapiens through eugenic laws nor the reshaping of the natural world through vast public works.

Applying scientific laws and quantitative measurement to most social problems would, high modernists believed, eliminate sterile debates once the "facts" were known. This lens-on-the-world has, build into it, a deeply embedded political agenda. There are, on this account, facts (usually numerical) that require no interpretation. Reliance on such facts should reduce the destructive play of narratives, sentiment, prejudices, habits, hyperbole, and emotion generally in public life. A cool,. clinical, and quantitative assessment would resolve disputes. Both the passions and the interests would be replaced by neutral, technical judgment.

High modernists aspired to minimize the distortions of subjectivity to achieve what Lorraine Daston has called "a-perspectival objectivity", a view-from-nowhere. The political order most compatible with this view was the disinterested, impersonal rule of a technically educated elite using its scientific knowledge to regulate human affairs. Partisan politics could only distort their wise decisions. Ultimately, rule on high-modernist lines would install a cosmopolitan ruling class that would, if they had them at all, set aside their nationalisms, their tastes, and their fears and perhaps even their scruples too, so as to govern scientifically. This aspiration was seen as a new

“civilizing project”. The reformist, cerebral Progressives in early 20th Century American politics were the closest thing, on the ground, to carriers of the high-modernist banner. Their gospel of efficiency, technical training, and engineering solutions implied a world directed by a trained, rational, and professional managerial elite.

How would this elite be created and selected? Here is precisely where high modernist claims are, or were originally, a radical departure from a history of privilege and entitlement. The idea of a “meritocracy” [Incidentally, the term “meritocracy” was coined in the late 1940s by the Englishmen Michael Young in his dystopian fantasy: “The Rise of the Meritocracy”, which mused on the disadvantages, for the working class, of a ruling elite chosen on the basis of IQ scores.] is the natural traveling companion of high modernism. No longer would a ruling class be the “accident” of noble birth, inherited wealth, or inherited status of any kind. Rulers would be selected, and hence legitimated, by virtue of their skills, intelligence, and demonstrated achievements. Meritocracy meant rule by the intelligent and knowledgeable. Here I pause to observe how other qualities one might plausibly want in positions of power: compassion, wisdom, courage, breadth of experience, drop out of this account entirely. Intelligence, by the standards of the time, was assumed by most of the educated public to be a measurable quality. Most assumed, furthermore, that intelligence was distributed, if not randomly, then at least far more widely than either wealth or title. The very idea of distributing, for the first time, position and life-chances on the basis of measurable merit was a breath of democratic fresh air. It promised for society as a whole what Napoleon’s promise of “careers open to talent” had promised the new professional middle class in France more than a century earlier. What seems new is both the breadth of the promise and, at the same time, the industrial Taylorism at its core: the dream of an efficient society achieving maximum production and progress because every bit of human material is deployed optimally to best utilize its capacity. This is the beginning of the vision that changes personnel departments into “Departments of Human Resources”. The rationalization Taylor had brought to the factory floor could now be applied to society as a whole.

Notions of a measurable meritocracy were democratic in still another sense. Not only did they supercede all claims to privilege by birth, they also severely curtailed the claims to discretionary power previously claimed by professional classes. Historically, the professions operated as trade guilds, setting their own standards, jealously guarding their professional secrets, and brooking no external scrutiny that would overrule their judgment. Lawyers, doctors, chartered accountants, engineers and professors were hired for their professional judgment—a judgment that was often ineffable and opaque.

How might claims to knowledge and expertise—claims that were not themselves transparent—be legitimated in a mass democracy? No longer was mere certification by the guild sufficient. Implying, as it necessarily does, that the citizen ought passively to defer to the mysterious decision of professionals and their guilds, the claim to expert knowledge is always suspect in a democracy. How then, to justify the professional privilege of “knowledge elites” who, increasingly, were responsible for “social engineering”, for structuring decisions that determined the life-chances of millions of their fellow citizens. Who will have a university education? Where will irrigation works, dams, and regional development funds be located. Who is eligible for scarce medical services and welfare benefits? It is in the conjunction of measurement, meritocracy, and democracy, I believe, that we can understand the runaway success [and later the troubles] of assessment techniques such as the SAT, the Research Assessment Exercises (RAE) in Britain, and Cost-Benefit Analysis.

Each technique is an attempt to substitute a transparent, mechanical, explicit, and usually numerical procedure of evaluation for the suspect practices of a professional elite. Each is a rich paradox from top to bottom. The technique is a response to political pressure; the desire of a clamorous public for procedures of decision and, in affect, rationing that are explicit, transparent and, hence, in principle accessible. Although each is a response to public political pressure- and here is one paradox—its success depends absolutely on appearing totally non-political: objective, non-partisan, even-handed, and palpably scientific. Beneath this appearance, of course, the technique is deeply political. Its politics are buried deep in the techniques of calculation: in what to measure in the first place, in how to measure it, in what scale to use, in conventions of

“discounting” and “commensuration”, in how observations are translated into numerical values, and in how these numerical values are used for decisions. While fending off charges of bias or favoritism, such techniques—and here is a second paradox—succeed brilliantly in entrenching a political agenda at the level of procedures and conventions of calculation that is doubly opaque and inaccessible.

The growth of testing regimens and cost-benefit analysis is not, then, a result of scientific progress, although they use the rhetoric of measurement and neo-classical economics as protective cover. Objectivity and mechanical rationality are a set of liberal political strategies designed to deal with political demands for equal opportunity and political transparency. They are apparently non-political ways of dealing with suspicion, distance, and volume. Thus, as Theodore Porter notes, ‘it is exactly wrong to assume that technocrats pursue objectivity except insofar as political pressures prevent them from doing so; instead, their mechanical objectivity arises as a response to political pressures to curb their discretionary power. When they are successful politically, the techniques of the SAT, Cost-Benefit Analysis, and, for that matter, its predecessor the Intelligence Quotient (IQ), appear as solid, objective and unquestionable as numbers for blood-pressure, thermometer-readings, cholesterol levels, and red blood-cell counts. The readings are perfectly impersonal and, so far as their interpretation is concerned, “the doctor knows best.”

The non-political, impersonal appearance of decisions by quantitative indices is the key to their political success. They seem to eliminate the capricious human element in decisions. Indeed, once the techniques with their deeply embedded and highly political assumptions are firmly in place, they do limit the discretion of officials. Charged with bias, the official can claim with some truth, that “I am just cranking the handle’-- of a non –political decision-making machine. The vital protective cover such anti-politics machines provide helps explain why their validity is of less concern than their standardization, precision, and impartiality. Even if the Social Science Citation Index does not measure the quality of a scholar’s work, even if the SAT doesn’t really measure intelligence or predict success in college, each constitutes an impartial, precise, public standard, a transparent set of rules and targets. When they succeed, they achieve the necessary alchemy of taking contentious and high-stakes battles for resources, life-chances, mega-project benefits, and status and transmuting them into technical, apolitical decisions presided over by officials whose neutrality is beyond reproach. The criteria for decisions are explicit, standardized, and known-in-advance. Discretion and politics are made to disappear by techniques that are, at bottom, completely saturated with discretionary choices and political assumptions now shielded effectively from public view.

The Scholastic Aptitude Test

The SAT, from its inception in 1926 has been the obsession—whether stumbling block or ticket to success—of American teenagers aspiring to higher education. Much has been written attacking it and much in its defense: most recently Nicholas Lemann’s fine volume “The Big Test: The Secret History of the American Meritocracy”. I will not repeat the story of the SAT’s rise to pedagogical hegemony. Lehmann, to whom I am indebted, tells it in great and convincing detail. Three aspects of that history, however, stand out for our purposes. First, the SAT is the lineal successor of various other tests: the Alfred Binet Intelligence Test (1905), Robert Yerkes”s Army IQ Tests during World War I, the Iowa Every-Pupil Test, the Stanford Achievement Tests, among others. It has become THE standard national test and a highly profitable commercial empire: The Educational Testing Service. Second, its great breakthrough to prominence came in 1959 when Clark Kerr, then Chancellor of the University of California system, decided to mandate SAT tests for all in-state applicants. Finally, the problems of validity, ‘colonization’ or ‘feedback’ effects, and the stifling of political debate were all fully understood by at least two ‘inventors’ of the test: Carl Bingham and Banesh Hoffman, later author of “The Tyranny of Testing”.

Whatever the reasons for the pre-eminence of the SAT in college admissions, the validity of the test as a measure of aptitude is not one of them. This issue, the basis of most critiques, need not long detain us. It is absolutely clear that aptitude or intelligence comes in many forms: aesthetic,

combinatorial, mechanical, metaphoric, imaginative, syllogistic, emotional, to name but a few. Correlations between such aptitudes, assuming each could be measured, would probably be insignificant. The SAT, however, encourages us to think of aptitude or acquired intelligence as a single, testable quality. Well….double quality actually: "Verbal and Quantitative", although, for many purposes it is the combined score that counts. The conceit behind the SAT is that a single, scalable metric can be devised to evaluate aptitude and that, for any individual, it is a stable index of scholastic aptitude. Even in this last, restricted sense, it fails its own test. SAT scores explain only 18% of the variance in first year college grades. Would we want to allocate one of life's great sweepstakes prizes on THIS basis without careful reflection and debate?

And yet, the SAT has become what Lehmann calls a "National Personnel Department," allocating a principal means to prestige, wealth, challenging work, and comfort in a race that is substantially done by age eighteen. It is arguably the key mechanism for reward distribution in our society and Lehmann rightly notes that it has never been politically debated, although it has been challenged in the courts.

Let us recall, as with high modernism and meritocracy generally, the democratic impulse at the origin of the SAT. Harvard President James Bryant Conant, patron of Henry Chauncey, founder of ETS was an unabashed proponent of meritocracy against considerable opposition, He believed that higher education should be reserved for a small, intellectual aristocracy (his word, borrowed from Jefferson) selected openly and democratically on the basis of mental aptitude tests. Not only did he believe that standardized tests could measure intellectual talent; he believed that too many students then benefiting from higher education were undeserving. He was out to unseat what Lehmann calls the "Episcopacy", or what one might call the WASPocracy. His plan would end the customary, cozy relationship whereby the Harvard (or Yale) admissions head would ask his friend at Andover or Groton or Saint Paul's to send a few of their "best boys". It would curtail the discretionary power of teachers' non-standardized grades and recommendations as well as the tastes and prejudices of admissions officers.

That prejudices might prevail behind a cloak of vague general standards is a documented fact. Jerome Karabel's close study of admissions at Princeton, Yale, Harvard, and Columbia from 1840 to 1940 is a striking case. To recapitulate it briefly, all four schools had, until roughly 1920, admitted the sons of alumni (most especially of wealthy contributors) who, they thought, could do the work. The rest of the class was filled out largely on the basis of school grades and tests that each school administered separately. Beginning in 1918, this system roughly doubled the percentage of Jews in the entering freshman class. Alumni complained and threatened to withhold support, while anxious administrators, even the few who were not themselves anti-Semitic, began to run for cover. By 1921, Harvard, Yale and Princeton collectively invented the concept of the "well-rounded applicant", the purpose of which I will leave to your imagination. Columbia, incidentally, being in the City, and having a somewhat different constituency did not adopt the "well-rounded" policy and became more heavily Jewish. By 1922, the percentage of Jews in the freshman class at the other schools had returned to their pre-1918 levels.

By contrast, the ideal of relying almost entirely on a standardized, non-discriminatory, open, and transparent aptitude test seemed a nearly utopian step away from a dark past of prejudice, patronage, and intrigue. It is, despite its manifold shortcomings, still the only common standard, the only metric gauge by which students from various backgrounds and school environments can be compared. Valid or not, it is the only common coin of the educational realm. Like a trade union norm, moreover, it provided an identifiable, public target for achievement. Students, their parents, and teachers all know more-or-less what it takes to succeed and, as we shall see, they take great pains to beat the system. Plagued with pathologies though it is, one would hesitate to abandon it altogether if the only alternative was the system it replaced.

Colonization:

I've grasped for metaphors that would convey how one small measurement- one that has little validity or predictive value—could have come to have such a huge impact. Malevolent viruses,

minute but murderous, come to mind. Let's just say that with respect to education, the SAT is not just the tail that wags the dog. It has reshaped the dog's breed, its appetite, its surroundings, and the lives of all those who care for it and feed it.

In emphasizing how deeply the SAT has transformed the world around it, I am telling a story that is both particular and generic. Were this not the United States and the DeVane Lectures, I could just as easily be telling a story about "exam hell" in Japan, France, Hong Kong, India, Brazil and many other countries where a single national exam (often more substantive than our own) are decisive in allocating careers and life-chances. Confining ourselves to national tests misses just how generic and modern this story is. It could be applied, with minor modifications, to almost any quantitative audit system deployed by powerful institutions to measure, by a single metric, a quality that is decisive for the distribution of resources, status, and opportunity.

Let's begin with the obvious. SAT scores are, in principle, like their ancestor the IQ, assumed to be stable (and hence uncoachable) for an individual. However, there's a man named Stanley Kaplan who's built a multi-million dollar business on the proposition that if you go to school with him, he'll improve your score. Many studies (not to mention his bank balance) have shown he is right. His flyers on campus cry out: "Get a Higher Score, Anytime, Anywhere" …."Kaplan Gets you In"---not just to college but into Law School, Business School, Medical School, Graduate School, Dental School via all the tests that the Educational Testing Service (ETS) has devised for these professional schools. He proclaims himself "The World Leader in Test Prep[aration]".

Kaplan is the tip of the iceberg. He saw his tutoring, from the 1930s on, as a 'poor man's private school,' giving outsiders, particularly Jews, a chance at advancement. Towards the base of the iceberg, there are thousands of cram courses and tutoring schemes working towards the same end. Stressing the generic quality of the feedback effect of the SAT, consider, for a moment, the larger effort in education to turn mere observations and decision criteria into targets. There are now thousands of educational consultants paid by anxious, wealthy families to secure their childrens' admission to anything from a quality kindergarten to an Ivy-League college. They teach three year olds to make "eye contact" and to demonstrate both leadership and sharing in kindergarten interviews! They not only recommend SAT prep courses, they steer their charges into volunteer work, hobbies, and summer activities that will fit the "profile" desired by a good college, help craft application essays, coach for interviews. Thus, the perfectly reasonable expectations of college admissions officials have worked their way back to colonize the life-world of aspiring teenagers in order to produce the simulacrum of what is required. I needn't remind you that money counts in this game; consultants and expensive prep courses are for the middle classes and higher. If this isn't the tail wagging the dog, I don't know what would be. Whole schools are now hiring firms like Achieva and the Princeton Review (aka ETS) to improve their school's record in college admissions by coaching.

Most demoralizing, perhaps, the SAT increasingly wags the curriculum dog in the schools themselves. Richard Atkinson, former Director of the National Science Foundation, an expert in testing, and now Chancellor of the University of California System recommended the California universities drop the SAT I, which tests abstract reasoning and substitute tests on specific subjects taught in high-school. Why? Because he was shocked to see twelve year olds drilled week after week in test-taking skills and sample questions at the expense of coursework itself. The term, now part of the language: "teaching to the test" expresses precisely that an observation with power behind it becomes a target and colonizes the very facts it is designed to record. To the degree that the SAT prevails, it encourages a more-or-less uniform national curriculum thereby stifling educational diversity and experimentation. "Teaching to the test" occurs nearly everywhere and we may expect it to flourish under President Bush's conception of school accountability via standardized tests. It thrived in Texas under Governor Bush's testing regimen, with students spending six weeks and more preparing for the dreaded TAAS (Texas Assessment of Academic Skills). Several districts (including Austin) have been caught cheating to improve scores; not surprising since districts, principals, and teachers can be rewarded or punished based on the results they achieve. Nor is Texas a deviant case. In Nassau County (Long Island?) N.Y., fourth graders spend half their time for two months in the fall on 'test prep' for a mandated reading exam and a like amount of time for two months in the spring preparing for the math and science exam.

In New York City, ex-School Chancellor Rudy Crew, who had removed five District Supervisors for flat or declining test scores, cited 32 schools (two principals, and 43 teachers) for cheating by helping students with the answers or simply providing pre-prepared answer forms.

The accountability-by-standardized-tests, born of anxiety, distrust, and quantitative management techniques has, then, produced a kind of rote, technique-driven Frankenstein in the classroom—often achieving the quantitative results required, but without the substance. Having devised a currency of quantity to represent the gold of qualitative achievement, the educational 'reformers' are repaid in fool's gold by an ironic alchemy. Much of the curriculum in elementary and high school has been "reverse-engineered" to raise scores on the SAT and other mandated tests. The Educational Testing Service promotes this retrofitting. It advertises curriculum programs tailored to the SAT: "Test Prep from the Test-Makers". First, they sell the American educational system a yardstick and then sell them the kit they will need to measure up.

Happily, since I signed up to give this lecture, a small but promising social movement of students and parents protesting 'teaching to the test" has blossomed. In western Massachusetts several hundred tenth-grade students boycotted the first of eleven days of tests to call attention to the effects of testing on their education. Two hundred Illinois students boycotted the new State English and Math Exams. Rallies against the effects of mandatory testing, often with the encouragement of teachers and parents, have been held in Louisiana, Florida, Wisconsin, and Ohio.

The 'colonization' or 'feedback' effect, however, doesn't end there; it goes far deeper. At one level, it is the business of the test makers [and the ETS as a private company dominating college admissions testing in unique in the Western world] to aggressively push their tests. An encrusted set of interests: test bureaucracies, text writers and publishers, curriculum developers, specialized teachers and counselors develops. And, in a perverse fashion, Stanley Kaplan and all the other 'test-beaters' are implicated too. They all feed the beast. The tests, though originally designed to assess individual abilities have spawned a new world of statistics and comparison, heavily freighted with rewards and punishments. Teachers, principals, schools, whole districts are judged by the average scores their students get—and woe to the administrator who cannot show improving scores. Little wonder that creative principals and supervisors have devised strategies to show apparent results, the same way a business can, by inventive accounting, produce a quarterly earnings statement that pleases investors, perhaps at the long run expense of the enterprise itself. In Texas, a Mexican-American group charged that schools were holding back minority ninth-graders who they feared would fail the tenth-grade exam. (NYT 10.30.01, p. a14) Many of those held back were discouraged and dropped out of school, thereby further improving the school's report card at the expense of the students'. Since the schools are also downgraded for 'dropouts', those leaving were classified as 'transfers'. Some schools have artificially inflated the numbers of learning-disabled students, whose scores are treated separately, to improve their standing. Finally, and most abjectly, quite a few states, finding that many of their students were failing the state-standardized tests, simply lowered the passing grade.

SAT scores have a flourishing life of their own outside the classroom. . High average SAT scores boost real-estate values; municipal bond rates are more advantageous in high SAT zones. SAT scores combine with other indicators to form the all-important 'scoreboard' ranking of colleges and universities by U.S. News and World Report. In turn, this generates pressures to select only applicants with high SAT scores so as to raise the institution's standing and perhaps pressure to encourages otherwise non-competitive applications so as to appear more selective.

A set of powerful quantitative observations, once again, create something of a social Heisenberg Principle, in which the scramble to make the grade utterly transforms the observational field. "Quantitative technologies work best", Porter reminds us, "if the world they aim to describe can be remade in their own image." I'm tempted to venture something of a law like generalization a la Parkinson. "Powerful [in the sense of influencing who gets what, when, and how] numerical representations of social practice also shape those practices; the two are necessarily connected by the feedback link of performance." [credit Rottenberg for the general idea]. It's a fancy way of

saying that the SAT has so reshaped education after its monochromatic image, that what it observes are largely the effects of what it has itself conjured up.

The SAT system has, over the past half century, been opening and closing possible futures for millions of students. It has helped fashion an elite. Little wonder that that elite looks favorably on the system that helped them get to the front of the pack. It is just open enough, transparent enough, and impartial enough to allow elites and non-elites to regard it as a fair national competition for advancement. More than wealth or birth ever could, it allows the winners to see their reward as merited and deserved, although the correlations between SAT scores and socio-economic status are enough to convince an impartial observer that this is no open door. The SAT, in effect, selected an elite that is more impartially chosen that its predecessor, more legitimate, and hence better situated to defend and reinforce the institution responsible for the 'naturalization of their excellence.'

In the meantime, our political life is impoverished. The hold of the SAT convinces many middle class whites that affirmative action is a stark choice between objective merit on the one hand, and rank favoritism on the other. We are deprived of a public dialogue about how educational opportunity ought to be allocated in a democratic and plural society. We are deprived of a debate about what qualities we might want in our elites, individually and collectively. We are deprived of a debate about how different skills might be taught and evaluated in out schools, insofar as curriculums simply echo the tunnel vision of the SAT. The SAT has become a threat to our education and to our democracy.