



ESSAY REVIEW

'The Big Men': a journalist's look at the Scholastic Aptitude Test

KIP TÉLLEZ

Our ordinary, human conceptual systems struggle daily with complex ideas and language. Indeed, the most complex thoughts often demand specific cognitive inventions to make their comprehension easier, and one of the most common strategies for improving comprehension is metaphor. Metaphor, of course, has a specific interpretation in literary studies, but, in the field of linguistics, its meaning is wider and generally considered a statement in which the unfamiliar is described in terms of the familiar. As Lakoff and Johnson note in their very influential book, *Metaphors We Live By* (1980), we are compelled to use metaphor to describe those concepts that elude simple description. For instance, our inability to describe the idea and purpose of human existence forces us to engage in rhetorical strategies that make this complex idea more real. How does one describe life? Our metaphorical understandings include common, pithy metaphors such as 'life is a journey', or 'life is a bowl of cherries'. But, metaphor can also be used to sharpen and deepen our thinking and shake us away from uncomplicated conceptual understandings. Lakoff and Johnson point out that the basis of all our experience is metaphorical. Indeed, Nietzsche referred to our understanding of the truth as a mobile army of metaphors.

A linguistic and conceptual tool closely related to metaphor is metonymy. Metonymy is sometimes considered a specific instance of metaphor in which a part of a concept or object stands for the whole. When referring to the monarchy, one might use the term 'crown', or when summarizing the viewpoints of workers, we might use the term 'labour'. US journalists often use this linguistic device when they report that the 'White House' has issued a statement or commented on policy. In this instance, the White

The book reviewed here is Nicholas Lemann, *The Big Test: The Secret History of American Meritocracy* (New York: Farrar, Straus and Giroux, 1999), ix + 406 pp., \$27.00 (hbk), ISBN 0-374-29984-6.

Kip Tellez is associate professor in the Education Department, University of California at Santa Cruz, 229 Crown College, 1156 High Street, Santa Cruz, CA 95064, USA; e-mail: ktellez@cats.ucsc.edu. His research interests centre on teacher education, assessment issues, and second-language education. He has published in journals such as *Journal of Teacher Education*, *Teaching and Teacher Education*, and *The High School Journal*. He is currently at work on a book examining teacher education and urban schooling.

House is metonymic for the President and the entire executive cabinet and staff. By using the term 'White House' as metonymy, the speaker is relieved of the tedious task of tracking the actual source and historical development of the policy, just as metaphors for the idea of life relieve us of the daunting task of facing up to the meaning of our existence. Both strategies give our conceptual systems a break, allowing us to simplify complex ideas, perhaps freeing up our cognitive processes so that we can focus on the consequences of an action taken rather than on who is responsible for a decision in the first place. Journalists, who are quick to show readers how a story will impact their lives, show a particular penchant for metonymy and use this speech act in ways that can both enlighten and deceive.

In a recent trade book entitled *The Big Test: The Secret History of American Meritocracy*, by the journalist Nicholas Lemann, metonymy is given full opportunity to flower as a journalistic invention that I believe serves to occlude general readers from the real workings of Lemann's subjects: the Educational Testing Service (ETS), generally, and the Scholastic Aptitude Test (SAT), specifically. Rather than exploring ETS as an immensely complex modern US bureaucracy built largely on technical expertise, Lemann grafts the ETS story onto the personalities of two men who helped ETS become a modern-day industry. This is journalistic metonymy of the highest order, simplifying complicated science and policy by focusing only on the personalities and drives of the individuals who 'led' the organization. In Lemann's drawn-out metonymy, he relies most on Henry Chauncey, the founder (in 1948) and first president of ETS, and William Turnbull, Chauncey's successor (in 1970) who oversaw ETS in its major growth period, to stand for the whole enterprise known as aptitude testing. Lemann contends that these men, together with James Bryant Conant, the former president of Harvard University, and Clark Kerr, the long-time president of the University of California system, were the architects of a new meritocratic order, a sifting and sorting of US intellectual talent in which SAT scores replaced wealth and privilege as the chief predictor of success. The personalities and intellectual passions of these four men serve as both the metonymic device and the narrative's structure to such an extent that I think *The Big Test* might have been more aptly entitled, 'The Big Men'.

In this review, I hope to demonstrate that Lemann's metonymy does little to help the general reader (the book is published by the US trade house, Farrar, Straus and Giroux, and is clearly written for an audience other than professional educators) understand the development, purpose and function of aptitude testing. Yet, in spite of this shortcoming, I suggest that *JCS*'s readers would benefit in several ways from reading Lemann's work. The book offers an 'outsider's' perspective on ETS and the SAT, while the possible use of selected passages and concepts might be useful in graduate courses in curriculum development or theory. In addition, I shall review the psychometric issues fundamental to understanding the SAT and suggest ways that Lemann and others interested in the SAT can help the general population comprehend the test itself. A better understanding of the technical properties of the SAT, I argue, may disincline American culture to have such a high regard for the test, and perhaps prevent the

hysteria that often surrounds the SAT and college and university admissions. Finally, I hope to convince readers that US curriculum specialists have an obligation to keep the SAT and other aptitude tests in their minds, if for no other reason than the SAT represents one of the most important determinants of post-secondary education participation and yet has almost nothing to do with curriculum. Indeed, I think it is fair to say that the more important the SAT becomes, the less important the school curriculum becomes.

Lemann's metonymy

Lemann begins his story in 1945, with Henry Chauncey sitting in church, appropriately enough, making the decision to leave his assistant dean position at Harvard to pursue his passion for mental testing. Henry Chauncey's life, as Lemann describes it, borders on both inspirational and desperate. Chauncey had grown up in the world of privilege that defines the expensive prep school track to the Harvard or Yale taken by members of the US *élite*. Aside from his freshman year at Ohio State, where he was, ironically, introduced to mental testing (Harvard in the early 1920s had no psychology department), the school tracking programme Chauncey experienced was the same one that for many years served to reinforce a well defined aristocracy in a nation that was not supposed to have an aristocracy at all. And Chauncey, having barely slipped into the *élite* track by using the position of his father and his superior athletic abilities, came to understand the world of privilege in a way that an outsider might be more inclined to: he found it ugly, replete with spoiled kids who wasted their education on parties and hazing. With the benefit of what amounts to Chauncey's memoirs of his time at ETS, a handwritten account of his early days of the SAT, and correspondence with Chauncey himself, Lemann then takes readers back to the early days when Chauncey wrote, in a 1932 letter to the educational researcher William S. Learned:

I am interested in the complete reorientation and reorganization of secondary schools aims and methods. . . . It is less complicated and more possible of fruition in the near future. It is concerned with the application of objective tests to college admission. . . . I myself should be only too glad to be of any help in such a venture, should it be projected (p. 20).

The words Chauncey uses to describe his now driving ambition, i.e. 'less complicated', and 'the application of objective tests to college admission', are emblematic of the way Chauncey confined his work at ETS and the general way he conceived of the project of testing for admissions. He never became a skilled psychometrician (although Lemann notes that he was highly skilled at describing to the general public what psychometricians did), nor did he have the same giant dreams of a societal reordering that James Bryant Conant and Clark Kerr had. Because Conant had grown up in the middle-class and had been accepted to a private academy based on a test score and not his pedigree, he, too, was an outsider to the Harvard establishment, even as its president, and he was much more appalled by the

old ways at Harvard than Chauncey. Conant was interested in nothing less than a complete rupturing of the existing social order of the nation. Lemann points out that in a short, unpublished book, *What We Are Fighting to Defend* (an unsurprising title, given that Conant wrote it in the early 1940s), Conant maintained the USA was run by an aristocracy based on heredity—a condition that put the entire nation at peril. Public education and the magic of aptitude testing, Conant believed, would create a society in which leaders were chosen on merit. Lemann, with his typically sharp eye for contradiction, writes:

Conant [with the writing of *What We Are Fighting to Defend*] had not abandoned his earlier ideas about education selection and sorting, so, in hindsight, the obvious question about the central idea of his unpublished book is: How can you build a classless society through the mechanism of relentlessly classifying the entire population? (p. 47).

Lemann notes that the stance Conant took in this book became his unquestioned goal for the remainder of his entire career. Conant, of course, became one of the most important shapers of the educational experience in the US, commenting on arenas well beyond admission policy. For example, it was Conant (1963: 1) who led the charge against formal teacher preparation, noting that, ‘I felt confident that I was an excellent teacher and I had developed my skill by experience, without benefit of professors of education’.

Clark Kerr, whose upbringing was even less privileged than Conant’s, oversaw (from 1958) the University of California system during the time of its greatest expansion and, as a member of the ETS Board of Directors, argued consistently for the use of the SAT to select elite students. Like Conant, Kerr saw advanced education, showered upon those who proved they deserved it based on grades and SAT scores, as the great sifter, the institution that would turn the staid aristocracy on its ear. However, with the election of Ronald Reagan as California’s Governor in 1966, which was followed quickly by the firing of Kerr and the dismantling of the impressive funding of his Master Plan for the state’s education system (if not the entire Plan itself), along with the passage of Proposition 13 (requiring state property taxes to be frozen), Kerr’s dream was largely over—but not before he convinced the University of California system to require the SAT.

The tension between creating a new elite and maintaining the notion that such an elite is both unnecessary and unwanted in a democracy, is the linchpin of the first few chapters of Lemann’s book, and he works the contradiction flawlessly. He points out that Conant, Kerr, and, to a lesser degree, Chauncey were not interested in expanding educational opportunities. In fact, even in the 1940s, they were concerned that too many students, most of whom they thought were mediocre, were entering higher education. Their goal was simple: to replace the aristocracy of wealth and privilege with a new one based on education and intellect—in other words, to put those like themselves in positions of power.

In addition to exploring the modern contradiction, Lemann helps us to see the historical context to the choosing of an aristocracy, by exploring a

little known dialogue in the USA between Thomas Jefferson and John Adams in the early 19th century. Jefferson was an early promoter of free public education in the USA. His ideas, although not entirely engaged during his lifetime, were picked up by Horace Mann, the educational reformer in Massachusetts, and others, who, in time, created the massive public school system now in place. Jefferson contended that an educational élite, or natural aristocracy, should be given charge of the nation's interest. Adams, in contrast, took exception to this position, pointing out that 'Your distinction between the aristoi and pseudo aristoi, will not help the matter. I would trust one as soon as the other with unlimited power' (Cappon 1959: 400).

Another player in Lemann's metonymy is William Turnbull, who served as second-in-command behind Chauncey for 26 years, then followed him as the president of ETS in 1970. In many ways, Turnbull, Lemann points out, was the antithesis of Chauncey. Chauncey was smooth, Turnbull was sharp; Chauncey was interested in human measurement of all kinds, Turnbull stuck to academic aptitude; and, finally, while Chauncey understood the politics of what ETS was trying to do, Turnbull was nearly apolitical at a time when politics mattered greatly. Lemann uses Turnbull's lack of experience in the political realm as a representation of the attack on nearly all sources of authority in the 1960s and 1970s. ETS, Lemann suggests, had become a part of the 'establishment'—a big secretive place where fates were decided without so much as a nod to public input or even justice. It was the perfect target, it would seem, for Ralph Nader, the consumer advocate, who in 1980 helped to publish a report by a heady undergraduate, Allan Nairn, and his colleagues. The report, entitled *The Reign of ETS: The Corporation that Makes up Minds* (1980), critiqued ETS in every imaginable way, from pointing out that ETS avoided even mentioning the contractual rights of ETS consumers, to showing how ETS officials were paid at roughly twice what a comparable position at a university would offer. Turnbull was simply not equipped to handle this kind of criticism. He could not understand why these bright young people, egged on by Ralph Nader of all people, would turn on the very institution he believed was responsible for their opportunity to hone their research and writing skills at a prestigious university. After a series of embarrassing bouts of public intoxication, Turnbull resigned in the early 1980s. ETS recovered, of course, with a new director, Gregory Anrig, at the helm. But, by this point in the book, Lemann has exhausted the potential of his Great Man metonym, and Anrig is barely mentioned.

By the time Turnbull was asked to resign the position of president, ETS had insinuated itself deeply enough into the culture of the USA to keep going with or without a president. It had co-opted the entire psychometric community, who by this time had so many mathematical tools and data to support the SAT that it was virtually guaranteed a long life. Indeed, it was the extensive arsenal of psychometric tools and the confidence they inspired that solidified SAT's hold on the imagination of the upwardly mobile within the USA. But, this is also the part of the story that Lemann's book ignores.

The contradiction of seeking to develop a 'new' élite in a nation of equals serves as the backdrop for the first few chapters of Lemann's book and it is this terrain he serves well. Along the way, we learn how ETS slipped its way into the national landscape through the gentle but persistent leadership of Chauncey, of the immense ego of Conant, the sociological fantasies of Kerr, and, finally, the precision of William Turnbull. But, apparently, Lemann wanted readers to know more about the SAT and how it has been used to eliminate people from opportunity without so much as a discussion. To this end, an out-of-place but necessary chapter entitled 'The Negro Problem' is sandwiched between a chapter on Henry Chauncey's son's experience as a student at Yale in the 1950s (which continues the metonymy, but fails to advance an understanding of the SAT or ETS) and a chapter on the fall of Clark Kerr. In that chapter, Lemann notes that in all this discussion of expanded college participation, none of the principals ever thought to consider that women or minorities might be considered for the new élite.

This chapter really deserves to be the focus of an entirely new book; it should also perhaps serve as a warning for what follows. In Chapter 20, with his metonymic analysis of ETS and SAT played out, Lemann abruptly begins a new story, and we follow Molly Munger, a Harvard-educated lawyer who led the fight against Proposition 209 (the legislation that ended affirmative action) in California in the mid-1990s. For nearly the remainder of the book, Lemann tries to catch readers up in Munger's work—in her intense desire to practice law in the interests of the oppressed, in her compassion for two African-American girls she hires as babysitters, and in her delicate humanity upon learning that Proposition 209 passed in 1996, in spite of her best efforts. I suppose Lemann's point in telling Munger's story is that sometimes the new privileged class of the educated and intelligent choose social justice over investment banking. But, these chapters do not hold much interest for someone who picked up the book to learn more about the 'Big Test' at ETS. And, the argument is not very compelling either, for, although it is true that the SAT gave Molly Munger a place at the élite university table, it did the same for Theodore Kaczynski, better known as the Unabomber, who was convicted in 1998 of killing three people and wounding several others by packaged bombs. In other words, ETS never set out to predict who would be kind and compassionate or even sane, and sternly refused ever to do so. For this reason, neither Munger's nor Kaczynski's actions deserve to be linked to the existence or the application of the SAT in US higher education.

Even after these weak, final chapters are considered, however, *The Big Test* is redeemed when Lemann returns to a discussion of merit and worth in the educational machinery of the USA and points out that schools should not simply be a great sorting machine. Lemann's work has clearly connected with the public (his book was the subject of a *Time* magazine cover-story and he has been interviewed on the US Public Broadcasting System 'Frontline' programme). Based on this exposure, it is arguable that Lemann's primary goal was to help the USA call into question, even discredit, the SAT and the élite group it has served. Lemann certainly does justice to that goal, but I do not believe he has helped the general audience

be less mystified by or obsessed with the SAT. In fact, I think he may have made the situation worse. What Lemann failed to do was to explain that the fuel for the SAT has not been the force of personality or an evil cabal of test item writers, but the development and eventual mastery of a set of statistical and psychometric tools so misunderstood by general audiences, and even educators, that ETS can now hum along with little interference. It is true that Lemann's book was not meant for psychometricians, and I am not suggesting that his metonymic, biographical approach is not more interesting than a technical understanding of the test, but I believe Lemann truly wants readers to understand why the SAT is so pernicious. And, to do that, he must provide readers with at least a general understanding of the mathematics behind test development.

Why avoid the mathematics?

I shall organize my critique around two themes: Lemann's failure to help readers understand the psychometric nature of the test, and his misunderstanding of the efficacy of SAT preparation courses.

The chief statistic ETS has used to support the use of the SAT is the Pearson Product-Moment Correlation Coefficient. This very simple statistic is at the heart of each and every analysis of data ETS uses to validate its aptitude tests. By computing a ratio between how two variables go together, or 'co-vary', ETS can move from stories of prediction to so-called mathematical precision in its work. The correlation coefficient is a very simple statistic to produce, requiring no more than simple algebra, and, more importantly, it does not imply causation. Lemann should have given more attention to explaining this statistic. Instead, he seems to encourage misunderstanding at several points in the text, including the following passage:

Most ETS tests settled into validities [*sic*], on a zero-to-one scale, in the 0.4 range. The predictive validity of grades was usually a little higher, and if you combined grades and [SAT] test scores (an exercise Henry Chauncey had invented when he was an assistant dean at Harvard), you got a higher predictive validity [*sic*] than from either one alone, somewhere around 0.5. That was certainly enough to make the tests useful, but still, all the SAT did by itself was explain ~15% of the variance in first-year grades in college, which was a pretty slender achievement, hardly commensurate with the magnificent role Chauncey had envisioned for testing (p. 86).

These statistics are the hard currency in the world of psychometrics generally and at ETS specifically. Lemann could have used this opportunity to explore what it meant by test validity, which, in this case, is a simple correlation. He, furthermore, could have helped the reader understand that the 'validities' can be understood as a value on a 'zero-to-one' scale, but that the scale is not linear, as his explanation implies. In other words, as the correlation between SAT scores and college grades goes down, the predictive value, or as statisticians optimistically call it, the coefficient of determination, will be disproportionately lower because its

value is computed by simply squaring the correlation coefficient. For instance, a correlation coefficient of 0.4 yields a coefficient of determination of 16%, understood as the percentage of variance that can be 'explained' by the relationship between the two variables. However, a correlation of 0.3 'explains' just 9% of the variance. So, a decrease of 0.1 in the correlation coefficient results in a much lower coefficient of determination. My point here is not to conduct an elementary lesson in statistical reasoning for *JCS* readers, but to suggest that Lemann should have done such a lesson for *his* readers. Like many others, I have taught these concepts to undergraduate pre-service teachers who both understand them and immediately see their value. Consequently, they understand the severe limitations of the SAT. And, when I see a few of them later as high school teachers in local schools, many share their frustration that even the guidance counsellors fail to understand these elementary concepts. For his part, Lemann misses a superb opportunity to examine the depth of this misunderstanding when he points out that Allan Nairn (the writer whom Ralph Nader helped produce the report critical of ETS) and colleagues did not have the expertise to produce r (the correlation coefficient), so the psychometricians ignored them. This point in the text could have served as a vehicle for helping readers understand the language of the scientists at ETS, which represents a language entirely incommensurate with the population at large. I did not expect Lemann to offer a Herrnstein and Murray's (1994: 593–623, 645–653) Bell Curve-like appendix, which served as a hocus-pocus side-show for Murray's attempts to reenergize the eugenics movement, but some attention to statistics would be welcomed.

At another point in the text, Lemann mentions what he incorrectly calls test equating, describing the way ETS develops new items. Naturally, ETS would much prefer to use the same items over and over. Item writing is difficult, demanding work. And item writers and the psychometricians who study test items come to regard them as trusted friends or hopeless losers. But, using the same items across testings would encourage cheating, or test pollution as psychometricians call it. So, in order to develop new items, test developers include non-scored items in each administration. Examinees do not know which of the items they are answering are non-scored or 'new items', so they answer them anyway.

Test developers then analyse the non-scored items to determine their difficulty, discrimination and their 'floor', the proportion of those who guess and get the item correct. The analysis and development of new items was made much easier and efficient with the invention of a statistical strategy known as Item Response Theory (IRT) (Lord 1980). And, although the algorithm for even a three-parameter IRT equation is not regarded as serious mathematics, the psychometricians were given a tool to make the task easier and obfuscate their work even further. Through the ceaseless application of the discrimination parameter of IRT, ETS has continuously improved the technical merits of the test while avoiding a discussion of initial validity. Thus, IRT allows psychometricians to use current test-takers to develop new items. For instance, ETS has on hand a huge bank of items for the SAT test, each with varying difficulty and discrimination. These items, which are considered valid, are compared

with the new, non-scored items. If an item behaves, based on IRT, like a previous item, it is considered a good item. In this way, ETS and other test developers are able to allow those who take the test to define the attribute they are measuring, thus avoiding the agonizing work of linking college coursework to items. The consequence of using IRT alone in developing new items is that the test simply selects over and over the same kind of intellectual capacity.

For its part, ETS in recent years has relied upon its principal research scientist, Howard Wainer, to defend the validity of the SAT. In a paper that attempts to explain why the SAT seems to lack validity for students attending the University of Hawaii, Wainer *et al.* (1993: 91) point out:

The SAT is the most widely used college admissions test in the US, with an annual volume of more than a million examinees. Its value has been widely debated, but empirically it is commonly found to predict first year college grades (FYG) about as well as high school grades. Since selection to university is based, to some extent, on a student's performance on the SAT, it is erroneous to calculate the correlation between SAT and FYG and call it *validity*.

And, in a footnote:

The correlation of the SAT with FYG is usually a little lower than that found with high school grade point average, but since the former is a score based upon a 2-hour exam and the latter on the aggregation of 3–4 years' performance in which a wide variety of school-related skills are included, it certainly seems remarkable to us that it can do as well as it does (p. 97, fn1).

It is difficult not to be suspicious of someone who speaks so confidently about his product. But, the psychometric point is that ETS may be moving away from defining validity in such narrow terms. In fact, I believe that ETS would be happier if, in discussions about the SAT, the notion of validity were avoided altogether, especially given the recent evidence suggesting that the correlation between the SAT and grades is weakening (Willingham *et al.* 1990). It would be consistent with ETS's historical position simply to provide scores to universities and make no claims about its validity. ETS has always tried to circumscribe carefully its tests, if for no other reason than to avoid costly and embarrassing lawsuits. To wit, Lemann notes that when Winton Manning, a former vice-president at ETS, tried to develop a system to account for social factors in the SAT score to remedy the enduring and troubling gap in scores by ethnic groups, Nancy Cole, the current president of ETS (appointed in 1994), cut off funding for his project. It is clear what Manning was trying to show: that success in college is related more to personal and social variables than SAT scores.

Many other researchers have set their sights on Manning's target. For instance, Wolfe and Johnson (1995) found that measures of personality, most notably a variable they call self-control, were better predictors of college grade point average than the SAT. Specifically, average grades in high school 'predicted' 19% of the variance in college grades, self-control was the next best predictor at 9%, and, finally, SAT entered in with an almost negligible 5%. But, in the end, I doubt whether even abandoning the

project of determining validity for the SAT would reduce the number of universities who require the SAT. I maintain that many college admissions officers recognize the limitations of the SAT but also believe it is necessary. In defence of their use of the SAT, they may point to grade inflation in high schools, overburdened university admissions offices, the need for so-called objectivity in admissions policies, and a general distrust of secondary school teachers.

In the arena of test validity, Lemann misses yet another opportunity to help general readers understand more about the world of testing. Because the SAT is an aptitude test, it differs in almost every way from achievement tests, the type with which nearly everyone is familiar. For instance, if secondary English teachers want to measure their students' knowledge of a novel, they begin by examining closely those instructional objectives they have used in teaching the book. Building test questions that directly relate to those ideas they taught in the novel, their test arguably measures what was taught, thus creating a valid achievement test. By contrast, an aptitude test has no instructional objectives; its validity lies only in its ability to predict future performance, a non-trivial psychometric challenge. The difference between aptitude and achievement tests is a discussion Lemann should have included in the book.

I believe Lemann's book will encourage readers to regard the SAT with even more mystery. Those whom we might think would know better continue to think of the SAT as an achievement test. It is not at all uncommon for middle- and upper-class parents to talk of their children 'studying for the SAT'. This is not a completely ridiculous idea to most people, based on the logic that if one does not do well on a test, it is because one did not study. But, when considering an intelligence or general cognitive aptitude test, one begins 'studying' for the SAT, an intelligence test for all intents and purposes, at the moment of birth and every moment thereafter. High school students who fail to meet their SAT expectations have little hope of improving their score by much. Lemann's book does little to help the non-educator public understand the difference between aptitude and achievement and the differences in how one might prepare for each. Lemann suggests at many points that anyone can improve their SAT score; all it takes is enough money to enroll in expensive (and private) test review programmes, like the suggestively named 'Princeton', or the Kaplan. On this score, Lemann pushes the general understanding of the SAT back 20 years.

Lemann is correct in reporting the ETS claims that the SAT is not vulnerable to coaching, but he would have done well by his readers to review the experimental effects of coaching programmes. Instead, the way in which he disregards ETS's 'no clear improvements' claim implies that ETS is the evil cabal we all suspected. But, the facts are clear on coaching, and many researchers, both inside and outside ETS, have come to the same conclusion. In a meta-analysis of coaching programmes, Becker (1990) found that on average SAT coaching programmes promise a nine-point increase for SAT-verbal and 19 points for SAT-math. These gains are hardly worthy of the time or money coaching programmes demand.

However, the minimal gains for students in general is not meant to suggest that certain groups, especially those who score below the mean on the first take, do not benefit from specific instruction or preparation. A study by Johnson and Wallace (1989) found that a coaching programme intervention for African-American low-income students did indeed show score gains on certain types of items from SAT-math. These items, not surprisingly, were those most related to the use of formulae typically taught in pre-collegiate mathematics courses. If students have not been introduced to the formulae and concepts on which the SAT-math is based, it stands to reason that by helping them to understand those essential concepts, their scores will improve. Of course, this type of 'test coaching' is remarkably similar to what we normally call teaching. It is also true that anyone who scores well below the mean score and retakes the test will likely do better on the second administration. But, I would not have expected Lemann to launch into an explanation of regression to the mean.

It is fascinating that more and more US high schools offer an entire semester of SAT preparation. So important is the SAT for their own rating—the state of Texas, for example, requires high schools to report the mean SAT scores in its accountability system—that educators appear increasingly willing to forego curriculum in favour of dreary test-strategy drills. More surprising is that high school students, who should be demanding relevance to their studies, accept these 'courses' as education.

Lemann should have treated his readers to a more complete discussion of the effects of coaching. With the wide audience he has reached (much wider than any curriculum theorist, I would argue), he could have encouraged readers not to waste their money, time and hopes on dubious coaching problems.

I have suggested that Lemann's attempts to tie the SAT and the bigger question of the meritocracy to the story of the men behind it fails to advance our understanding. But, the SAT remains, at its heart, a set of questions to be answered (new on each administration, thanks to non-scored items). Lemann is right in that Chauncey was not a psychometrician, but ETS is no longer driven by the musings of dreamers hoping to re-order society. Instead, it has become a very comfortable home for statisticians and psychometricians whose expertise or even interest is not political philosophy. ETS is now a fully developed factory whose products are regarded as trusted precision tools in the otherwise messy, messy world of education. As evidence, ETS's new growth market is the Test of English as a Foreign Language (TOEFL), an achievement (not aptitude) test required of most international students hoping to attend US universities. The TOEFL is a clear winner for ETS: an achievement test for which items are easy to write, an eager and willing examinee population who generally lacks the sophistication to challenge the test or ETS, a test with a clear need, and one whose costs are easily passed on to the students, not the universities.

By failing to address the technical aspects of the development of cognitive aptitude tests, and by neglecting to assist his general readership in adopting a more sceptical view of the SAT and test preparation,

Lemann's book has helped to hand over educational studies to statisticians. In this error, however, he is not alone. Educational research in the past 20 years has moved further and further away from considering what actually happens in schools and classrooms and the important questions we should be asking about those contexts, and instead has become increasingly a branch of applied mathematics.

The field of economics provides a good example of this point. Early in the 20th century, economists working in the academic sphere largely studied how people and institutions made hard choices about allocating their resources. Included in this discussion were questions about the moral and ethical dimensions of their work and an attempt to create a larger axiological and epistemological framework for economics. But, at some point about mid-century, economics transformed itself from a true social science into a branch of applied mathematics, in which many of the leaders of the discipline did not discuss applied issues at all. Rather, they became expert in using large sets of data (made possible by the use of the emerging computer) to make predictions about future economic behaviour. And, these predictions, in turn, were made possible only by the use of multi-variate statistical analyses, a set of equations for correlating multiple variables that allows for the researcher to assess the effects of an individual variable in relation to many others.

A cursory look now at the field of economics reveals a nearly complete transformation. For instance, in the latest volume of *Journal of Economic Theory*, every article tested a mathematical model. I contend that education is only ~20 years behind economics in its transformation from a field that takes seriously the underlying social and psychological consequences of its work and instead becomes the domain of applied mathematicians who know little about teaching and learning but who, nevertheless, are asked to render judgement on educational—and cultural and political—matters. Clearly, the blame for this condition cannot be laid at the door of *The Big Test*. All who aspire to place the discussion of educational aims in a philosophical context must share responsibility when the technicians take over. And, I believe that in order to prevent education's fall into the purely technical realm that economics now finds itself, we must help ourselves and those around us understand how mathematics can both inform and confuse. Lemann's primary fault was his difficulty in seeing that a conspiracy in testing is really nothing more than highly technical people carrying out an ill-defined mission.

Conclusion

In spite of shortcomings, *The Big Test* is timely. We Americans are so immersed in an era of accountability and quantification that the culture has failed to articulate the purpose of higher test scores in the first place. The point of higher test scores, in today's logic, appears to be higher test scores, which is a striking example of begging the question, and Lemann tries to help us understand that the SAT fails to measure most of what we value in our leaders.

Lemann leaves his best work for the 'Afterword', in which he returns to the ideas that served him well in the first half of the book, and at one point he invites us to think about the role of college by noting, 'It's a given by now that it matters a great deal *whether* you went to college; it doesn't, however, have to be supremely important *where* you went to college' (p. 350). Although I deeply appreciate the topic Lemann raises, I think that, on this score, he could not be more of a dreamer. Even if we succeeded in upending the SAT as the great dispenser of educational opportunity (in the form of admission to the élite universities) in the USA, the existing aristocracy, chosen by the SAT, would find another way to stay in power. The élite universities in the country were opened to a new kind of student with only a marginal struggle from the old aristocracy. But, this former aristocracy, whose privilege was based on inherited wealth, was losing power anyway. A progressive income tax structure, inheritance taxes, the Depression of the 1930s, and a new, even wealthier group who 'earned' their fortunes, all contributed to the collapse of the insular Harvard/Yale crowd and the funhouse they created at their universities. And, Lemann's primary theme is that this group has been replaced with a new aristocracy who have been chosen not by birthright but by SAT scores. I can promise that this new aristocracy will not relinquish power so easily. Remember that this aristocracy has 'merit' on its side—it 'earned' its privilege. It believes, like James Bryant Conant, that they *deserve* their position, and that people who share their analytical qualities deserve to join them. The old aristocracy may have had its faults, but it never had 'science' and 'mathematics' so convincingly on its side. We now have an aristocracy based not on bloodlines but on an abstract ability to see patterns and relationships in word and number puzzles on a timed paper-and-pencil test. Lemann invites us to wonder whether this new élite is serving us any better.

I shall conclude by noting that the metonymic account offered in *The Big Test* fails, because the SAT's importance came about as a result of hundreds of factors, including advances in test scoring, computers and psychometric sophistication, and a distrust of high school curriculum. The men whom Lemann chooses to profile only serve as figureheads of this larger movement. And, in describing the effects of the SAT on the culture, Lemann limits his analysis to the effects of intelligence and schooling at the level of the individual. This analysis, in my view, is wrong. In its place, we can view those who are able to solve the pattern and relationship puzzles as part of a larger social milieu, a view promoted in *Liberalism and Social Action* (Dewey 1935). Dewey argued that intelligence could not be defined at the individual level, suggesting that 'liberalism has to assume the responsibility for making it clear that intelligence is a social asset and is clothed with a function as public as its origin, in the concrete, in social cooperation' (p. 67). In a subsequent section, he notes that development of the world's new technologies of his day (e.g. mass steel production) were not the result of individual invention, and that a survey of recent scientific breakthroughs 'brings home the actual social character of intelligence as it actually develops and makes its way' (pp. 68–69). People use their individual intelligence in a distributed fashion. The problem with the SAT is

that working with your neighbour will get you thrown out of the examination room.

References

- BECKER, B. J. (1990) Coaching for the scholastic aptitude test: further synthesis and appraisal. *Review of Educational Research*, 60(3), 373–417.
- CAPPON, L. J. (ed.) (1959) *The Adams-Jefferson Letters: The Complete Correspondence between Thomas Jefferson and Abigail and John Adams: Vol. II, 1812–1826* (Chapel Hill, NC: University of North Carolina Press).
- CONANT, J. B. (1963) *The Education of American Teachers* (New York: McGraw-Hill).
- DEWEY, J. (1935) *Liberalism and Social Action* (New York: G. P. Putman's Sons).
- HERRNSTEIN, R. J. and MURRAY, C. (1994) *The Bell Curve: Intelligence and Class Structure in American Life* (New York: The Free Press).
- JOHNSON, S. T. and WALLACE, M. B. (1989) Characteristics of SAT quantitative items showing improvement after coaching among Black students from low-income families: an exploratory study. *Journal of Educational Measurement*, 26(2), 133–145.
- LAKOFF, G. and JOHNSON, M. (1980) *Metaphors We Live By* (Chicago: University of Chicago Press).
- LORD, F. M. (1980) *Applications of Item Response Theory to Practical Testing Problems* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- NAIRN, A. (1980) *The Reign of ETS: The Corporation that Makes up Minds* (Washington, DC: Ralph Nader).
- WAINER, H., SAKA, T. and DONOGHUE, J. R. (1993) The validity of the SAT at the University of Hawaii: a riddle wrapped in an enigma. *Educational Evaluation and Policy Analysis*, 15(1), 91–98.
- WILLINGHAM, W., LEWIS, C., MORGAN, R. and RAMIST, L. (1990) *Predicting College Grades: An Analysis of Institutional Trends Over Two Decades* (Princeton, NJ: Educational Testing Service).
- WOLFE, R. N. and JOHNSON, S. D. (1995) Personality as a predictor of college performance. *Educational and Psychological Measurement*, 55(2), 177–185.