Enemies of Promise: Why America Needs the SAT

William C. Dowling

I.

In March 1998 UCLA Chancellor Albert Carnesale, a cab driver’s son who had earlier risen through faculty ranks to become Provost of Harvard, announced to California taxpayers what he thought was a bit of good news. UCLA had just enrolled its most academically outstanding class in history, with freshmen SAT scores putting it among the most selective public universities in the nation. The response, from UCLA’s Affirmative Action Coalition, African Student Union, and other campus groups, was outraged, leading to “days of action” at which Mr. Carnesale was accused of wanting to turn UCLA into the “Harvard of the West.” These protesters wanted to concentrate not on SAT scores but on a different set of statistics they saw as being an intolerable consequence of Proposition 209, the constitutional amendment that had abolished many affirmative action policies. Since 1997, the number of black, Hispanic, and American Indian freshmen admitted to UCLA had fallen from 2,066 to 1,327, a decline of 36 percent.

On the face of it, the opposition between Chancellor Carnesale and his opponents looks like a simple battle of statistics. Yet for critics of the SAT, the verbal and mathematical reasoning test annually administered to over a million high school seniors by the Educational Testing Service (ETS), the real truth about the confrontation lay in their conviction that Carnesale’s statistics constituted an unrecognized public fraud. To the Affirmative Action Coalition, citing a drop in minority enrollments is simply to use a convenient numerical idiom to talk about real 18-year-olds born into poverty or culturally deprived circumstances, actual minority youngsters being denied their rightful place in a public university system. To cite SAT scores is, on the other hand, to parade numbers that are quite literally meaningless, a sort of statistical mumbo jumbo devised by a dominant social class to keep the less fortunate in their place.

Anyone familiar with the rhetoric of FairTest, the Cambridge, Massachusetts organization that has led the national assault on the SATs, will be aware that I am not exaggerating for effect. FairTest’s widely-circulated “SAT Fact

William C. Dowling is a professor of English at Rutgers University. He is author most recently of The Senses of the Text: Intensional Semantics and Literary Theory (University of Nebraska Press, 1999). Please address correspondence to Academic Questions / NAS, 221 Witherspoon Street, Second Floor, Princeton, NJ 08542-3213; editor@aq.nas.org.
Sheet,” for instance, darkly warns readers that the SAT is “a direct descendant of the racist anti-immigrant Army Mental Tests of the 1920s.” An only slightly less inflammatory rhetoric is used by John Katzman, one of the founders of Princeton Review, an SAT-coaching outfit that makes no secret of its contempt for the test. “It’s an arbitrary, biased, somewhat pointless exam,” says Katzman, “that doesn’t test anything important.”1 But criticism of the SAT is not restricted to a few strident critics of ETS and standardized testing. Here is Larry Stedman, a respected researcher on tests and educational policy: “The SAT is a speed endurance test, made up of 200 problems in 3 hours, or more than one problem per minute. And much of it is made up of what can, at best, be charitably called verbal conundrums and math puzzles.”2

Yet there is a paradox here. Whatever the ultimate verdict on the SAT, its recent role as a scapegoat in the Affirmative Action debate turns the story of the test’s actual origins inside out. For the authorizing presence behind national-level standardized testing in America was James Bryant Conant, who, having in 1933 risen to the presidency of Harvard from humble origins, was determined to open the university to intellectually promising students from throughout the United States, regardless of social or economic background. The SAT was developed—largely by Charles Chauncey, then an assistant dean at Harvard under Conant—as part of the Harvard National Scholarship program, through which Conant hoped to break the hold on Harvard of a wealthy, eastern, prep-school-educated elite, and with it a pernicious mind-set including (as Nicholas Lemann says) a “casual and unearned assumption of superiority” and an “inability to see immigrants, Catholics, Jews, and the poor as fully human.”3

My own interest in the SAT controversy arose initially from my experience as an English professor at Rutgers University—particularly as a regular teacher of English 219, a course in the “close reading” of poetry—and then, as I did more research, from a growing sense that several hidden elements in the controversy have never been gotten across by ETS or others sincerely wishing to defend the test in terms of Conant’s original vision, as a common yardstick permitting bright youngsters from disparate backgrounds to aspire to top colleges and universities. (I should say that everything that follows concerns the test’s verbal section, normally abbreviated as SATV. This test, which is an improved version of the original SAT, has borne the burden of opponents’ attacks. The more specialized mathematical section was added only belatedly, in an attempt to improve the SAT’s predictive power.)

I began to suspect that the SATV measures something real and important several years ago, when one of my Rutgers students came by my office to ask for a recommendation to English graduate programs at Harvard, Yale, Berkeley, and Chicago. At this point, like most English professors, I was fluent in the idiom of SAT verbal scores—and scores on the Graduate Record Examination (GRE), the next-higher-level version of the SAT—and, like university profes-
sors generally, had always been disposed to discount them more as badges of socio-economic status than a measure of verbal aptitude or cognitive ability. This particular student, who was from a working-class New Jersey family, had done brilliant work for me as a freshman in English 219 three years before. As is customary, she had brought me her grade transcript—it showed all A’s—but, being abnormally conscientious, she had also brought a complete packet including her high school SAT and recent GRE scores. In both, she had scored 800 on the verbal section.

As anyone familiar with the ETS scoring system will be aware—the scale on both the SAT and GRE runs from 200 points at the bottom to 800 points at the top—these were spectacular scores. Since this student had done superb work in an English course concentrating on such difficult poets as Donne and Shakespeare and Milton, I began to wonder if the SATV, which I had so often heard debunked as a measure of verbal or cognitive ability, was meaningless after all. When a departmental task shortly thereafter gave me an opportunity to compare the grades of my English 219 students over several years with their incoming SATV scores, I compiled a simple statistical chart. What I found was that the SATV scores had an extraordinarily high correlation with final grades, and that neither, in the many cases where I had come to know my students’ personal backgrounds, seemed to correlate very well with socio-economic status.

Nonetheless, I make no claims as a statistician. I mention these correlations only because they led me to several years’ research into the SAT controversy. My point in what follows, based on a great deal of reading about the SAT and its implications for educational policy, plus an extensive analysis of actual SAT tests, will simply be that the intuitive sense of most high school and university teachers that some students are verbally brighter than others, and that this bears in a direct and crucial way on their ability to do college-level work, seems to me to match up very well with verbal and cognitive levels as measured by the SATV. The way to understand why this is so, I have come to believe, is not to argue about scores and demographic statistics, but to go inside an actual SAT verbal test to try to get a feel for how it works. That is what I will undertake to do below. Before doing so, however, I want to address one important preliminary issue, which is the mistaken or misleading notion that the SAT, if it is to have any validity as a test of cognitive ability, should be a successful predictor of college grades. For this is a point on which ETS has done as much as opponents of the SAT to encourage widespread confusion.

II.

Almost every attack on the SAT begins by asserting that ETS has fraudulently promoted the test as a predictor of student performance. As usual, FairTest leads the chorus. “What is the SAT Used For?” asks FairTest’s “fact sheet.” Answer: “The SAT is validated for just one purpose: predicting first-
year college grades. It does not do even this very well. Testmakers acknowledge that high school grade-point average (GPA) or class rank are the best predictors of first-year grades, despite the huge variation among high schools and courses." ETS has never been able to counter such charges satisfactorily, not least because in the early days of the SAT it did, in fact, promote the test as a grade predictor. The result has been endless puzzlement about what the SAT is and means.

To understand why the originators of the SAT were disposed to think of the test as a grade predictor, it helps to recall Conant's vision of a common yardstick in relation to the Harvard National Scholarships. Suppose, for instance, that it is 1935. You are a bright high school student from a rural community in Iowa or Oregon, to which the rumor of Harvard has barely penetrated. No one in your family has ever gone to college. The education given by your grammar school and high school is rudimentary. But you have read your way through the local public library—the amount you read is a source of constant amazement to your parents and neighbors—and gotten good grades in your classes. I, on the other hand, simply represent the latest generation of my wealthy Massachusetts family to attend Groton, where, more through sheer tenacity than any outstanding intellectual gifts, I have also managed to compile a good grade point average.

When the SATs appeared on the national scene, there was an element of romance and even suspense in this scenario, a combination of what might be called its prince-and-the-pauper and gunfight-at-the-OK-Corral aspects, that appealed to those who, like James Bryant Conant, had risen from obscure origins through sheer intellectual ability. For if, in the scenario we are imagining, you and I walk into geographically distant rooms on the same day to answer an identical set of questions, and you emerge with a 780 verbal score to my 610, it seems entirely reasonable to suppose that your endless hours of reading have served better to educate you than my six years at Groton. The notion of the SAT as a predictor of grades was originally meant simply as a validation of this point. Were you and I to enter Harvard in the same freshman class, you would have the capacity to outperform me academically—that is, in once-conventional terms, to get better grades—despite the incontestable superiority of my Groton education.

Today, as is admitted by virtually everyone except FairTest, this sort of rationale has been blown out of the water by runaway grade inflation at both the high school and college levels. In some high schools, it is now not unusual for more than a third of the senior class to graduate with an A average, and the percentages of A's given out at colleges and universities are endlessly reported in the national press: Georgetown, 42 percent, Harvard, Yale, and Williams all over 35 percent, most public universities 30 percent or higher. Nor do the grades given out even at elite institutions seem to have much relation to student ability or performance. As reported by his lawyer, an imposter who in the
early 1990s gained admission to Yale with a fake transcript earned a B average in the two years before he was caught, even though his GPA at the community college he previously attended had been only 2.1. In this climate, predicting grades is a little like predicting the price of potatoes during the Weimar inflation of the early 1920s, a pointless and erratic enterprise.

Still, there is a more important reason why the notion of the SAT as a grade predictor has always been meaningless. It turns on the distinction, still taught in introductory logic courses, between necessary and sufficient conditions. Breathing oxygen is, for instance, a necessary but not a sufficient condition of your remaining alive over the next six months. To specify sufficient conditions, one would have to list all the other things—food, water, a certain minimum body temperature—that permit you to function as a physical organism. This is why the notion of the SAT as a predictor of grades made no sense even in the days when grades bore a genuine relation to student performance. For grades involve not just intellectual ability but such factors as discipline, mental concentration, and efficient use of time. In no era could a freshman who chose to get drunk every night rather than to study be predicted to do well in his English classes, whether or not he had entered college with a 780 SAT verbal.

In reality, a high SATV score has always been a measure not of whether a student would perform well in college courses demanding verbal ability, but only of a sine qua non—that is, a necessary condition—for doing so. The other factors explain why admissions officers at selective colleges and universities spend so many hours poring over supplementary materials for clues to what in a more innocent age could be called personal and moral character—letters from teachers, employers, and counselors, reports from alumni interviews, the winning of a Westinghouse Science Award or a national violin competition. A 780 verbal assures, in short, that this student can do the work, even if the college is as competitive as Williams or Harvard. But only the other materials in the packet are able to suggest whether this is also a student who, under the pressures of an Ivy League freshman year, might wind up throwing in the towel. On these matters the SAT necessarily remains silent, nor has anyone ever really pretended otherwise.

ETS has never chosen to dwell on this obvious point, one suspects, because in America the very notion of the SAT as a sine qua non is a political powder keg. We find it easy enough to admit that a student with a 780 SAT verbal score is not going to do well at Harvard if he spends his evenings getting drunk. It is very difficult to face the fact that a student with a 480 SAT verbal, no matter how disciplined or hard working, simply could not do well there. The popularity of professional sports in American society, it is sometimes said, is due to their being the one domain in which we have somehow come to terms with the idea of natural ability. No American male is offended at being told that, no matter how hard he practiced, he could never be Michael Jordan or, indeed, a starting player in the NBA. Yet the thought that many students with
good study habits and high aspirations do not have what it takes to do the work at Harvard or Berkeley remains a source of deep discomfort. It is ETS’s quite understandable reluctance to come out and say this that has left it open to attack from organizations like FairTest.

III.

The greatest source of confusion about the SATV is that it purports to measure something—"developed verbal ability," as it is called in the psychometric literature—that can be made to look mysterious, to the point that, like powers of augury or mental telepathy, one may be led to doubt that it exists at all. Yet verbal aptitude isn’t mysterious. If you were able to read the first sentence of this paragraph with perfect comprehension, for instance, you were demonstrating a high level of what ETS means by developed verbal ability. Partly, this just means that your vocabulary is large enough to let you understand the English used by educated speakers. (Consider, in that first sentence, purports, psychometric, augury, telepathy.) But it also means that you’ve done enough reading to be able to follow complex sentences, as a violinist or pianist, after years of lessons and practice, is able to sight-read a musical score. That’s what people who use the SAT to evaluate student aptitude mean by “developed” verbal ability.

Now take a moment to think about what was happening as you were reading the preceding paragraph. One feature of your comprehension is especially helpful in understanding how the SATV works. Notice that, although the paragraph consisted of several sentences, you had to understand those sentences not simply as isolated combinations of words, but as adding up to a whole greater than its parts. This is to say that higher-level reading comprehension includes a set of logical, syntactic, and semantic relations so complex that modern linguistic theory has barely begun to see how they work. This is especially true in the area of “pragmatics,” which deals with the rules of relevance by which we get from one sentence to the next even in ordinary conversation. The most important thing the SATV tests is the student’s grasp of these relations.

To see how it does so, let’s look at the verbal sections of an actual SAT, taken by students nationwide in November 1996. (One of the major consequences of the so-called truth in testing law passed by the state of New York in 1979 is that ETS now releases “used” SATs for students who want to gain some familiarity with the test before taking it. This test is taken from 10 Real SATs, published by the College Board in 1997.) Like every other SATV, this one poses questions in three categories: sentence-completion items, in which the test-taker is asked to fill in blanks from a list of words or word pairs; analogy items, which test a grasp of logical relationships through matched word pairs; and reading comprehension items, which ask students to read one or more passages of prose and then to answer a series of analytic questions. The format is standard: the
first two verbal sections, containing items of each kind, are timed to take 30 minutes. The final section, reading comprehension only, is timed at 15 minutes.

Let’s look at each category in turn. The point of the SAT’s sentence-completion questions is to measure vocabulary and powers of syntactic analysis. To see how they do so, consider two sentences with blanks representing the missing words. I am going to make them extremely simple—though no simpler than many items ETS designates as difficulty level 1—to give the clearest possible idea of the principle involved:

a) Seeing that his right ______ had come unlaced, George bent down and laced it back up.

b) Feeling intensely ______ after her hot afternoon working in the garden, Susan went to the refrigerator and got a cold ______.

At this very elementary level, the SATV sentence-completion format suggests why critics like Larry Stedman have been led to say that it involves little more than verbal conundrums. Who could not see that a word like shoe is needed to complete (a), or that only some pair like thirsty . . . drink will do in (b)? The point that tends to get lost in discussions of the SATV is that, with verbally competent test-takers, the format works in just the same way up to a very high level of complexity. That is, a student with high verbal aptitude will treat every item in the sentence-completion sections in the way readers on all levels treat items like (a) and (b), seeing what words are needed to complete the blanks even before looking down to examine the available choices offered.

Let’s see how the same principle works on the sample SATV I’ve chosen. The first item in the test is, as it happens, a sentence completion. Let me begin by giving the sentence in isolation:

1) Many cultural historians believe that language has a ______ purpose: it serves not only as a means of communication but as a means of defining culture.

This is what the SAT calls a difficulty level 1 item—since I’m going to be talking about difficulty levels quite a bit in what follows, I’ll abbreviate this as dl from now on—and, like the elementary examples I offered a moment ago, it suggests in very clear terms what sort of word is needed. To answer (a) above, we surmised that whatever George bent down towards had to be something with laces that could come untied, hence shoe. Here, the part of the sentence following the blank makes it obvious that the word we need is going to have to specify two purposes that language can serve, namely (1) as a means of communication, and (2) as a means of defining culture. So what is needed is an adjective meaning “two.” Here are the answers offered by the test:

(A) foreign   (B) literary   (C) false   (D) dual   (E) direct
The answer, of course, is (D). But it's essential to see, before dismissing items like this as too elementary to warrant discussion, that students at a low level of reading comprehension will find even this question difficult. For the sentence in which the blank occurs contains vocabulary—cultural, communication, defining—that will be over the heads of some test-takers, and the syntax that permits one to see that an adjective meaning “two” is necessary will have been impenetrable to many remedial-level students. Suppose, for instance, that a test taker has grasped that a purpose is being talked about, but has not understood that the sentence then goes on to specify exactly two purposes. Suddenly, other answers on the list seem plausible. We are talking about language, after all, and language can be used with a literary purpose (writing a poem), or a direct purpose (ordering someone to perform an action), or a false purpose (lying to a grand jury). Where some students see a single answer standing out with crystalline clarity, in short, many others will see a bewildering thicket of possibilities.

With the same principle in mind, let's look at a two-blank item. As before, I'll give the sentence first, letting you project possible answers before looking at the actual list of choices:

4) Born ____, children will follow their natural inclination to explore their surroundings with a ____ that belies the random appearance of their play.

The first blank, one sees, permits only a very narrow range of possibilities. The word, whatever it turns out to be, will have to be an adjective modifying children, and this adjective will have to describe the natural inclination of children to explore their surroundings. The key to the second blank, just as obviously, is the term belies, which means to give the lie to or be inconsistent with—e.g., “George's outgoing manner belies his actual shyness.” Here, what is belied is the apparent randomness of children's play: they look like they're playing randomly, but what they're really doing, says the sentence, is exploring their surroundings. Here are the choices offered the test-taker:

(A) innocent . . . deviousness
(B) serious . . . merriment
(C) curious . . . purposefulness
(D) eager . . . purposefulness
(E) aware . . . casualness

The answer is (C). This is a difficult item, which means that, although most students taking this particular SATV got it right, some got it wrong. Nor is it hard to guess what reduced those students to puzzlement: a test-taker who doesn't know what belies means also won't see the logic that leads to (C). At higher levels of difficulty—dl4 and dl5, which is the range, roughly speaking,
in which students who get over 700 on the SATV operate comfortably—it is the relation of vocabulary to reading comprehension that matters most. Still, it is the *relation*, and not merely the vocabulary, that SAT sentence-completion items test at this more advanced level. Consider, for instance, this *d* question:

9. The traditional process of producing an oil painting requires so many steps that it seems ______ to artists who prefer to work quickly.

Notice that, as in earlier examples, it is still the syntax of the sentence that is doing the work. Painting in oil *requires many steps*, so that it is demanding, laborious, time-consuming. And *artists who prefer to work quickly* don’t want to spend this kind of time. To such artists, therefore, the process of producing an oil painting would inevitably seem ____. Now in this case, we can’t predict exactly what is going to go in the blank, but a reader with moderately good verbal comprehension will nonetheless have seen readily enough what sort of meaning is wanted: *tedious, pointlessly time-consuming, so tedious that it seems like it will never end,* etc. So we look at the list of choices offered by the test:

(A) provocative  (B) consummate  (C) interminable  (D) facile  (E) prolific

The answer is (C), obviously, but this will not be obvious to someone who has no idea what *interminable* means. To such a person, by the same token, the other choices will look like a welter of wholly opaque terms: someone who can’t recognize *interminable* when it occurs in a sentence is not going to have much better luck with *provocative, consummate, facile, or prolific*. Critics very often try, when talking about the way the SAT measures vocabulary, to imply that it involves little more than an arbitrary word-list of useless terms. But in fact the test operates on principles that seem entirely reasonable. Someone able to describe his wait at the doctor’s office as *interminable* is also likely to know what you mean if you describe a dress or a remark as *provocative*. Someone who has no idea what *interminable* means, on the other hand, is going to find a great deal of intellectually demanding writing impenetrable.

In a moment I want to look at the SAT’s analogy and reading comprehension sections, but let us pause a moment to take stock. From David Owen’s *None of the Above: Behind the Myth of Scholastic Aptitude*, published in 1985, to the latest fulminations of FairTest, critics of the SAT have worked hard to suggest that the test’s verbal portion consists of pointless conundrums followed by arbitrary and meaningless lists of answers from which the hapless student is compelled to choose while the clock ticks remorselessly away in the background. Yet (critics say) there is no particular logic involved in choosing “right” answers. They may be answered “correctly” only if one is privy to a secret code issued by ETS to students lucky enough to be born into wealthy families.
Even the few actual SATV questions we've looked at, however, will suggest a far different picture. For in every case it has clearly not been ETS but the sentence that is choosing one item from the ABCDE list. More importantly, the way in which it does so exactly mirrors a process central to reading comprehension at any but the most elementary level. The same, as we've seen, goes for the vocabulary demanded by $d14$ and $d15$ questions: the way to grasp the meaning of *interminable* or *provocative* or *consummate* is not to be born into a wealthy family but to get to the public library and begin reading. So far, in short, the SAT may simply be seen to be testing the cognitive powers that come into play when one is reading Aristotle or Thoreau or Henry James. On a test meant to measure students' readiness for college-level reading and thinking, this does not seem unreasonable.

To get a line on the SATV's analogy questions, it's only necessary to see that college-level reading involves a grasp of logical as well as syntactic relations. At the simplest level, the analogies test elementary part-whole and part-part relations of a sort that would have been familiar to a schoolchild in ancient Greece. At a somewhat higher level, the SAT's analogy sections also test a complex sort of logical relation that derives from semantic entailment, the precise nature of which is a matter of considerable debate in contemporary linguistic theory. But this sort of entailment is something every competent speaker of a language deals with intuitively and, in ordinary cases, unerringly. Thus, for instance, any native speaker of English will agree that the sentence *George had a nightmare* must "logically imply" *George had a dream*. To see why and how such entailments work, it is necessary only to subject them to what Kant called the law of contradiction. Suppose someone told you—meaning it in absolutely literal terms—that *Harry killed Sally*, but then failed to understand that Sally was dead. You would have every right to assume that this speaker doesn't wholly grasp the meaning of at least some English sentences.

When they simply test such matters as the part-whole relation, the SATV's analogy items tend to be pretty elementary, with any difficulty coming not from the logical relation but from the vocabulary. Here, for instance, is a $d13$ question from our test's analogy section:

**Preamble: Statute**

Notice that, as with the sentence-completion items we looked at earlier, the question even in isolation sets up a purely formal expectation: since a *preamble* is something that comes before a *statute* or law to explain its meaning or purpose, whatever satisfies the analogy is going to have to display the "coming before" and also the "declaring purpose" features of a preamble. We could even think up, before looking down to see what choices are offered, various pairs that satisfied the analogy: *introduction: manual*, for instance, or *prologue: play*, either of which gives us written texts in a relation to each other analogous to *preamble: statute*. Here is the list of choices actually offered:
(A) interlude: musical  
(B) conclusion: argument  
(C) foreword: novel  
(D) epilogue: address  
(E) premier: performance

The answer is, of course, (C). The difficulty that makes this a d55 item must be coming, clearly, from the vocabulary: a student who does not know what an epilogue is will not know that it comes after a preceding text or speech or performance, and therefore cannot be seen as analogous to a preamble. In the same way, a test-taker who has no idea that an interlude is something that comes in the middle of a performance will be at a loss about what is supposed to be analogous to what in this case. (Such examples suggest why, as has been widely reported in recent years, students who take Latin in high school tend to improve their SAT scores. Even a first-year Latin student will know that the pre-of preamble signals “beforeness” and the inter-of interlude “in-betweenness.”)

To see how the SATV tests students’ grasp of semantic entailment, let’s look at an item that comes at the very top (d55) of its difficulty range. Semantic entailment, once again, involves seeing that the meanings of certain words are so to speak “built in” to the sense structure of others: to know that Peter is a bachelor must, to a competent speaker of English, be to know that Peter is unmarried. This is the sort of logical implication that comes into play in items like the following:

Querulous: Complain

Here, as before, a great part of the “difficulty” is coming from the vocabulary: there are quite literally millions of students enrolled in American colleges and universities who could not tell you that a querulous person is one who complaining a lot. But to a test-taker who does know that, the item will scarcely be more challenging than preamble:statute. A querulous person will be seen to complain a lot, one would want to say, as a untruthful person will be seen to lie a lot, or a glutinous person will be seen to eat a lot. Here are the actual choices given for this item:

(A) silent:talk  
(B) humorous:laugh  
(C) dangerous:risk  
(D) deceitful:cheat  
(E) gracious:accept

The answer, which in this instance is the sort of thing that gets you into an Ivy League college, is (D).
To get the basic idea of the SATV’s reading comprehension sections, it’s necessary only to look at a sample of the prose the test-taker is asked to read. The final section of the SAT we’ve been examining, for instance, is entirely reading comprehension: two passages followed by 12 questions and timed to take 15 minutes. Both passages are about the effects of TV news, with passage 1 arguing that it stimulates critical thinking because the viewer gets a direct view of events—the writer’s example is the way images of civil rights sit-ins made Americans outside the South aware of segregation—and passage 2 arguing, to the contrary, that television trivializes events and the whole conception of public awareness or public duty, by presenting the news simply as entertainment. Passage 2 consists of four paragraphs. Here is the first paragraph and the beginning of the second:

“Now . . . this” is a phrase commonly used on television newscasts to indicate that what one has just heard or seen has no relevance to what one is about to hear or see, or possibly to anything one is ever likely to hear or see. The phrase acknowledges that the world as mapped by television news has no order or meaning and is not to be taken seriously. No earthquake is so devastating, no political blunder so costly, that it cannot be erased from our minds by a newscaster saying, “Now . . . this.” Interrupted by commercials, presented by newscasters with celebrity status, and advertised like any other product, television newscasts transmit news without context, without consequences, without values, and therefore without essential seriousness; in short, news as pure entertainment. The resulting trivialization of information leaves television viewers well entertained, but not well informed or well prepared to respond to events.

The species of information created by television is, in fact, “disinformation.” Disinformation does not mean false information, but misleading information—misplaced, irrelevant, fragmented, or superficial information—that creates the illusion of knowing something, but that actually leads one away from any true understanding.

Whenever I discuss the SATV with students who got scores in the 700+ range, the reading comprehension items tend to elicit an incredulous grin. (“They tell you,” one of my students said, “that the chicken crossed the road, then they ask ‘What crossed the road?’ and ‘What did it cross?’”) Yet when we recall that the SATV is designed to tell whether students who take the test are intellectually prepared for college, and that the ability to read at a certain level is more essential to college-level learning than any other factor, the reading comprehension sections make a good deal of sense. The way to grasp the point of items in this category, in short, is not to think about the small minority of students who find them embarrassingly easy, but about the hundreds of thousands who find them insuperably difficult.

Three questions from our sample SATV, all based on the passage quoted above, will show how the reading comprehension section works. The first asks whether the test-taker has understood a key word in the argument:
6. The word “mapped” in line 4 most nearly means ________

Since understanding mapped in line 4 simply involves seeing that it is a metaphor—i.e., that the writer is not talking about actual maps or geography here, but the way TV news defines social or cultural reality for its viewers—it’s not easy to see where a wrong answer could come from. There are two possibilities, however, and they are precisely the ones that students with lower-level reading ability will tend to seize on. The first is that mapped is literal, having to do with maps and globes and measurement. In the list of answers below, (A), (B), and (C) play on this possibility. The second is that mapped does indeed mean something metaphorical, but the test-taker’s vocabulary doesn’t include one or more of the words given as choices, as might happen with (D) and (E) below. Here are the choices:

(A) plotted on a chart (B) planned in detail (C) measured (D) defined (E) verified

The answer is (D).

In the second item, the specialized meaning of a key term prompts the question. Even some students with good verbal ability will not be exactly clear on the difference between disinformation and misinformation, and, in fact, the writer is putting something of a spin even on the normal meaning of disinformation to drive home a point. So the question asks specifically how the writer is using the term:

8. According to Passage 2, the “disinformation” mentioned in line 13 affects television viewers by _________________.

In this case, the wrong answers don’t demand a lot of discussion. They are wrong either because they’re nothing more than random or irrelevant associations with things mentioned in the passage, or because they directly contradict something in the passage, as when (A) gives the test-taker the option of filling in the blank with “leading them to act on false information,” while the writer of the passage has clearly said that “disinformation does not mean false information”;

(A) leading them to act on false information
(B) causing them to become skeptical about television news
(C) giving them the mistaken impression that they are knowledgeable
(D) making them susceptible to the commercials that accompany the news
(E) turning them against certain political leaders

The answer is, of course, (C).

Finally, the SATV reading comprehension section includes a sort of summary item meant to see whether or not the test-taker has gotten the essential
point of the passage as a whole or some major part of it. In principle, since any reasonably complex argument might be summarized in a variety of ways, these are the items that could most plausibly be attacked by critics hostile to the SAT. Yet in practice, the SATV minimizes ambiguity through the simple measure of listing four absurdly wrong possibilities along with one obviously right one. Here, for instance, is a question in this category:

6. According to Passage 2, television news is presented in a manner that serves to ____________________________.

Look again at the passage above, taking a moment to notice that the writer says that TV presents news “as pure entertainment,” and that newscasts “leave television viewers well entertained, but not well informed.” (Later on, in the part of the passage I haven’t quoted, the writer says TV news “is packaged as entertainment,” and that its viewers have been “entertained into indifference.”) Here are the choices:

(A) hold leaders accountable for their policies
(B) entertain viewers
(C) define lies as truth
(E) make complex issues accessible
(D) exaggerate minor political blunders

The answer, as you may have guessed, is (B).

There, in a nutshell, is the SATV, the object of FairTest’s wrath and the reservations of other, less strident critics. As I’ve tried to show, looking in detail at any actual SAT is likely to leave one wondering what the fuss is about. My own guess is that opposition to the SATV arises from two distinct sources. One is an instinctive dislike of “standardized” testing, with its implications of rigidity and impersonality, plus a suspicion that the SATV doesn’t really measure anything important. Critics in this category could be satisfied, presumably, by being shown that the test does measure something real, and that its standardization, as with, say, the standard use of Olympic distances in athletic events held at different times and different locations, serves an obviously equitable purpose. The other category consists of critics who, despite the contrary assertions they so vociferously make for public consumption, are wholly aware that the SATV does measure verbal and cognitive ability, and oppose it on precisely that ground. I want to end by looking at critics in this category.

IV.

In the overcharged political atmosphere surrounding the SATV in recent years, with controversy focusing on such issues as the disparity between African American and white scores or the relation of SAT percentile to family
income level, it has been easy to lose sight of an otherwise simple fact, which is that any student who has spent a great deal of time reading is virtually assured of a high SATV grade. In *A Is For Admission: The Insider’s Guide to Getting into the Ivy League and Other Top Colleges*, in my opinion the best book on college admissions currently available, Michele A. Hernández, a former associate dean of admissions at Dartmouth College, returns to this point again and again. “It is not surprising,” observes Hernández, “that the students who get 800 on the verbal SAT scores are always the ones who were read to a lot and then developed a real love of reading as they grew up. Even if they were not challenged in school, these children could read in their spare time and thus learn many of the skills necessary to succeed in college.”

The same observation, which is entirely borne out by my own twenty-five-year experience in the university classroom, bears on another controversial issue, the effect of SAT coaching on the scores of those, mainly students from the upper middle class, who can afford test preparation courses. Here the lines have been clearly drawn, with people like Princeton Review’s Katzman claiming substantial score increases from coaching, ETS maintaining that the gains are, on average, insubstantial. What almost never gets mentioned is Hernández’s point that gains on the SATV, whatever their size, come mostly in the middle range: “It is not uncommon . . . to raise a verbal score from 450 to 600 or from 570 to 680. What is almost impossible is to jump into the 720 to 800 range, even if you are starting in the high 600s. With a few exceptions, the students who score over 740 or so are simply voracious readers, students who have been reading seriously since they were very young and have continued to do so all their lives.”

One major reason that SAT coaching is able to yield such results in the middle range has to do with what experts call test familiarity: doing well not by paying attention to what is being tested, but how the test works. Princeton Review, for instance, has from the first concentrated on teaching its clients to “beat the test.” An example described with evident relish by David Owen in *None of the Above*, for instance, consisted of a trick to be used on the SAT math test. Instead of doing mathematical calculations on geometry problems, Princeton Review students were taught to tear off a corner of their test booklet—a perfect 90° angle—and use it as a protractor, eliminating wrong choices on the answer list by eye measurement of the scaled diagram provided with the problem. On the SATV, students who don’t know the answer to a question are taught various techniques to eliminate “obviously” wrong answers. Precisely as with card-counting in casino blackjack, this does effectively increase the probability of guessing right choices over an entire run of test items.

This sort of test-beating instruction has become less effective over the last few years, as ETS has worked to close the loopholes spotted by outfits like Princeton Review. But even when it was successful, such coaching might be viewed as amounting essentially to training in petty dishonesty, as though an
Olympic coach, having found a loop on the course unsupervised by race officials, were to teach marathoners to turn in “faster” times by taking a short cut. (One suspects that this is why Katzman is so strident in his denunciations of the SAT: it’s only by encouraging an utter cynicism about the test that Princeton Review’s test-beating strategies can be made to look anything other than what some people might consider a bit shady.) Yet even Princeton Review gestures in the direction of substantive teaching, as with its famous “hit parade” of vocabulary items, based on a frequency analysis of past SATs and taught to its clients as preparation for the sentence completion and analogy sections.

Our analysis of a sample SATV shows readily enough why this sort of thing will yield at least limited results. A sentence completion item we looked at earlier, for instance, required interminable as a correct answer. Even a student who knew the meaning of interminable purely through rote memorization would have a substantial advantage over a test-taker who had no idea what it meant. Among SAT coaches who carry this sort of substantive approach even farther, insisting that students spend an entire year learning the meaning of vocabulary words in context, the gains are even more dramatic. But then it is no longer clear that we are dealing with SAT coaching rather than simple, old fashioned, effective teaching. “People don’t back their way into a good score,” says Arun Alagappan, founder of Advantage Testing, a hugely successful tutoring operation that demands a great deal of its teenage clients. “Those who do well are conscientious students.”8 Not surprisingly, Alagappan has a positive view of the SAT as a measure of cognitive ability.

Nonetheless, Michele Hernández’s point that coaching gains come primarily in the middle range will leave even most of Advantage Testing’s clients below the 750 SAT level, above which scores tend to measure a lifelong devotion to reading. “Every time I have seen a student with an 800 verbal score,” reports Hernández, “there has been confirmation throughout the application that the student is a reader.” “Teachers mention it,” she adds, and “the student often talks about loving literature from a young age.”9 Critics of the SAT sometimes manage to imply that large numbers of worthy students are being turned away from Harvard and Yale each year due to low scores, but this is not the case. As ETS likes to point out, over 90 percent of SAT-takers nationwide are admitted to the college of their first choice. In effect, this means elite colleges and universities draw their entering classes almost entirely from the small cadre of devoted readers Hernández describes. For everyone else, the test is largely a formality.

For colleges and universities themselves, however, a great deal hangs on SATV levels. A point almost never discussed by either side in the SAT controversy, for instance, no doubt because it is as much a political powder keg as the issue of natural cognitive ability, is that there is an average SATV level below which no institution can sustain a college-level curriculum, for the simple reason that such a curriculum will consist of materials beyond the comprehen-
sion of a student body with lower-level SATV scores. For any such curriculum will necessarily include not only writers like Aristotle, Thoreau, and Henry James, but history or economics or anthropology textbooks assuming college-level reading ability. My own analysis, based on difficulty levels of SATV items in relation to materials taught at Rutgers, suggests that a 580 SAT marks the lower limit of college-level reading comprehension, a combined 1130 SAT the minimum needed for college-level work.

Yet many colleges and universities have mean or average SAT scores below this level, which is no doubt why one so often hears the complaint that many public institutions in America, operating essentially on a policy of open admissions, have become little more than glorified high schools. Statistics suggest that such pessimism is not unwarranted. Nationwide, points out Edwin S. Rubenstein in a recent survey, approximately 30 percent of incoming freshman are today placed in at least one remedial course. At state universities remediation rates are even higher—35 percent, 48 percent, and 39 percent in New York, Kentucky, and Georgia, respectively.10 At a further extreme, nonselective admissions can actually produce a situation in which students read and write at a grade-school level, which no doubt explains the note of impatience in the voice of the mayor of New York as he called recently for an end to open admissions at CUNY: “Open enrollment is a mistake. . . . By eliminating any meaningful standards of admission and continually defining down standards for continuation, the entire meaning and value of a college education has been put in jeopardy for the many who are ready, willing, and able to meet and exceed higher standards.”11

In recent years, the competition to get into Ivy League and a few other selective colleges has been treated in the national press as little more than a frenzied pursuit of status by the middle class, a hysteria as irrational as the rush of lemmings to the sea. The competition is, indeed, ferocious, as witness Hernández’s account of a recent Dartmouth class: “The average combined score for the admitted class of 2000 at Dartmouth was over 1410. In fact, for the class of 2000, the average scores of all 11,400 applicants who applied to Dartmouth (this includes all the weakest applicants in the pool) was 662V, 677M, almost 1340 combined.”12 Yet to see this as perfectly rational behavior, it is necessary only to see that, for the roughly 10,000 students who got turned down, the application process was not only a struggle to get into Dartmouth, but also to not be compelled to attend a school where most students are incapable of meeting college level demands. This is another point that almost never gets mentioned in the national SAT controversy.

Very bright students tend to think about getting into Harvard or Amherst or Dartmouth in the same way as a gifted violinist might hope someday to make the Chicago Symphony Orchestra or a gifted basketball player the NBA. Looking at such organizations from the outside, what one tends to hear about is their prestige. To those inside, however, prestige is almost always far less
important than a chance to exercise natural ability or talent in a setting where others are equally talented. This serves to explain why demographic categories, so strongly depended on by critics of the SAT, are almost meaningless at this level. As it is impossible to predict in advance into what household a musical genius or NBA star will be born, it is virtually impossible to pinpoint where the sort of avid or insatiable reader described by Hernández will emerge.

At the highest level of performance, demographics can do no more than suggest the conditions under which talent or ability is likely to emerge, which in the cases we’ve been considering would include the degree to which a family or a culture emphasizes reading, music, or sports. This is the point, for instance, of the example of the National Basketball Association that is sometimes invoked in this connection. Though millions of young American males aspire to the celebrity and wealth that comes with making the NBA—the average salary of an NBA player last year was $1.2 million—African Americans as a group are tremendously overrepresented while other groups (Jews, Hispanics) are underrepresented to the point of invisibility. Still, no one would waste a moment trying to argue that NBA selection is somehow biased in favor of African Americans and against Jews and Hispanics, if only because the need of professional teams to win guarantees that slots will be awarded to the best players. Yet a similar demographic disproportion at the higher levels of SAT performance is the entire basis of FairTest’s argument that the test is culturally biased.

Nor is their argument without serious consequences for American society. For to take it seriously is to justify the sort of policies advocated by Derek Bok and William Bowen in The Shape of the River; or by Harvard president Neil Rudenstine in several widely reported speeches, which seek to correct demographic disproportion through brute force, rigging the admissions process at selective institutions to achieve desired results in racial and ethnic categories. It is against this background that it is salutary to recall the vision that originally motivated James Bryant Conant when he set out to open the university up to talented youngsters from across the United States, without regard to family income or educational background. The essence of Conant’s notion of policy, especially as it involved the Harvard National Scholarships and the then newly-introduced SAT, was that it viewed America not in terms of demographic categories but of talented individuals.

Conant understood, in short, that factors like ethnicity and family income operate only as loose determinants of intellectual achievement. Nor is the point less valid today than during Conant’s presidency of Harvard. Rich families and white families, we may suppose, will go on turning out their usual substantial quota of average and below-average offspring, children entirely happy to grow up in a world of MTV and skateboarding and total intellectual vacuity. Millions of other children, no matter what their socioeconomic background, will choose to grow up in that same world. Even in a country that
grants everyone twelve years of free education—a situation undreamt of in past centuries—few will want to read books. But scattered here and there in American society will be children who read their way through the public library. For them, as Conant foresaw, the SAT is less a test than a means of becoming visible to a world outside their own locality.

As Conant understood, the ultimate stakes have to do not with educational policy but with civic purpose. The symbolic value of the SAT lay, for him, in its power to remind an American society corrupted by the worship of wealth that its moral basis lay in rewarding talent or ability wherever found. This is why he instructed the Harvard National Scholarship project to cast so wide a net for what in Conant’s generation were called promising youths. In a series of articles published in the Atlantic Monthly in the 1940s, Conant called this appeal to democratic or egalitarian principles a “new radicalism,” by which he meant a return to the roots of American experience. Only against the background of such earlier idealism, perhaps, do we glimpse the sense in which FairTest and other groups who today demand abolition of the SAT would until very recently have been seen as, to borrow a famous phrase coined in Conant’s own day, the enemies of promise.

Notes
