

Hail to Thee, Ol' Dart U: Retake Policies and the SAT

by Jesse Fuchs
www.numbertwopencil.net
Updated 2/24/03

I.

At the American Prospect website, there's a recent article by Duke assistant professor Jacob L. Vigdor. The article, entitled "College Try: Why universities should stop encouraging applicants to take the SAT over and over again," is about the general college admissions practice of selecting only the highest SAT scores from applicants. He's against it; he thinks that the policy fosters class and race inequality, and that a system in which retaking the test is less encouraged would help close that gap with no downside worth mentioning. It's possible that he's right, and I'll deal with that possibility starting with part VII. But first things first; Vigdor's whole argument is wrapped in a piece of rhetoric so baldly, ridiculously manipulative that it needs to be debunked before I can begin to address any of the real issues it points towards. To begin his essay, Vigdor asks the reader to imagine the admissions office at a hypothetical Academy of Dart-Throwing:

This type of institution would clearly wish to accept only those applicants showing the highest potential as professional dart players. It would therefore be reasonable to expect such schools to use a dart-throwing "exam" as an admissions criterion.

I'd never make this analogy myself, for reasons I'll get into at the end of this essay, but I'll accept it for the moment as a reasonable abstract model. It's with Vigdor's description of the exam, though, that his analogy starts spinning out of control:

Now suppose that the dart-throwing academy employed the following method of assessing ability: Each applicant arrives at a dart-throwing center early on a Saturday morning, pays a \$25 fee and, after a three-hour wait, throws exactly one dart at the dartboard. Those who place their darts closest to the center of the board are considered superior applicants. Of course, because there is a certain amount of chance involved in where the dart lands, this method will imperfectly assess ability. Some very good dart players will have a bad throw; some poor dart players will have luck on their side.

Now, waiting for three hours to throw a single dart is a terrible way to separate the dart-sheep from the dart-goats, and any Academy of Darts that relied such an admissions test would quickly lose its accreditation and/or liquor license. (Unless we're talking about a CIA Dart-Throwing Ninja Sniper program, in which case this process might make perfect sense, but I'm going to assume that this isn't the case.) A real Academy of Darts would need some sort of admissions test, to be sure, but I imagine they'd develop a much better exam than Vigdor suggests. Instead of making you wait three hours to participate in a five-second contest, for instance, they'd use the entire three hours for the test. Instead of throwing a single dart, the applicant would participate in 5 or 6 rounds, each testing different dart-throwing skills. They'd throw 25 or 30 darts in each round, each of which would count a little bit towards their total score. And instead of automatically charging everyone \$25, the testing service would provide waivers for students who couldn't afford the fee, even if they were taking the test for the second time. In other words, they would probably construct a test that sort of resembled the SAT. Let me put this bluntly: any student who actually trusts in Vigdor's analogy and spends the first three hours of the SAT waiting around, only to attempt to answer a single question on the test absolutely perfectly, is not going to get a favorable score no matter how lucky they are.

II.

The analogy wouldn't be that bad if it was just another case of a journalistic hook that's catchier than it is precise. But what it does is far more disingenuous: by framing SAT scores as though determined by a single throw of a dart, Vigdor misleads the reader into thinking that, by taking the test multiple times, a rich but inept SAT-taker is virtually bound to eventually get a lucky throw in. Consider the following stunning assertion:

Given a sufficient number of Saturday mornings and sufficient funds to continue paying the testing fee, even the least-skilled dart player could be assured of a favorable admissions ranking.

Let's take this, for the moment, at face value. The least-skilled SAT student, by definition, would be one who knew nothing, and thus guessed at random on every question. This is actually being slightly generous; such a student, given average luck, would get 26 (out of 138) questions right for a combined score of 480. To get below that, you have to do worse than random. (All scores in this article are taken from the May 2000 SAT, which I've randomly picked as the scale to use for all the following calculations — other tests may vary by 20 or 30 points in either direction). Let's also be generous at the other end, and lowball a "favorable admissions ranking" at a combined score of 1100 — which, in dart terms, is somewhere around hitting the black-and-white circle surrounding the little circle that surrounds the actual bullseye. This score, split evenly between Math and Verbal, would require the student to get 87 questions right. What are the odds that, by dint of 60 or so lucky guesses, the student will stumble blindly onto it?

I can't tell you, but it's not my fault; my TI-83 only does scientific notation to so many places. Lowering the target score by 100 points at a time, I managed to finally hit a calculable probability at a combined score of 800, which is below the bare minimum for Division I academic eligibility. The chances of this student getting the 53 questions right needed for this score are just under 1 in 2 million. There are plenty of words I could use to describe these odds, but 'assured' is assuredly not one of them.

III.

Vigdor's clearly guilty of massive hyperbole here, but underneath it is a point worth addressing: what about more realistic scenarios, where a student who's already doing decently takes the test multiple times in hopes of getting some lucky breaks? Here, Vigdor's argument still isn't as strong as he'd like you to believe, but three things do make it significantly stronger:

1) The higher scores get, the more small differences matter, especially at the most competitive schools: the practical difference between a 480 and an 800 is probably less than that between a 1280 and a 1400.

2) The closer you get to the top of the scale, the more one question can change a score. In the middle of the SAT scale, the bell-curve distribution causes scores to repeat: if you have a 500 in math and you get one more question right, you may still end up with a 500, whereas in the 700s, one question right or wrong can mean a difference of 20 (or, in some instances, as much as 40) points.

3) Paradoxically, (and as long as the student guesses at all,) the fewer guesses they make, the more likely they are to (in relative terms) get lucky — i.e., the closer they will come to resembling Vigdor's single dart throw. Here's a thought experiment to illustrate that last one. Let's say that you can enter a contest in which you win \$1000 by accomplishing one of two tasks: flipping a coin five times and getting all heads, or flipping a coin 100 times and having it land heads 60 or more times. Which is the better bet to take?

Going 60-40 may seem easier than going 5-0, but it actually requires more luck; it only occurs 2.8% of the time, while getting five out of five occurs 3.1% of the time, making the less tedious bet also the safer one. It's the Iron Law of Casinos: it's temptingly easy to beat the odds on two or three rolls of the die or spins of the wheel, but if they get you to stick around long enough, even a 1% advantage on the House's part will inevitably catch up with you.

IV.

So, with all that in mind, let's look at four more hypothetical students, whom I'll call Shermmy, Linus, Marcie, and Joe Cool:

Shermy is an impossibly average examinee. He gets the first half of each section completely right and blindly guesses on the tougher half, giving him, with average luck, 82 right and a combined score of 1040: 520 Math, 520 Verbal. (All subsequent total scores discussed split within 10 points of each other.)

Linus gets the first 3/5ths of each section completely right. Linus is also able to make educated guesses on the harder questions — he eliminates, on average, two answer choices from each one, giving him a 1 in 3 chance of getting them right. With average luck, this nets him 100 right and a combined score of 1180.

Marcie aces the first 2/3rds of each section and can eliminate three answer choices on the rest, giving her a 50/50 chance of getting them right. With average luck, she gets 114 right and a total score of 1320.

And Joe Cool is, of course, an SAT whiz — he gets all but the last 4 (or 2, in the half-length sections) questions absolutely right, and has a 50/50 chance on the few remaining. With average luck, he gets 127 right and a combined score of 1460.

So, the big question is: what are the chances that any of these students will, through lucky guessing, achieve a score significantly higher than their “true” score on an SAT? Thanks to the magic of binomial probability distributions, here's those odds for your delectation:

	+ 50	+100
<i>Shermy (1040)</i>	8.5%	0.1%
<i>Linus (1180)</i>	15%	0.3%
<i>Marcie (1320)</i>	6%	0.3%
<i>Joe Cool (1460)</i>	6%	2%

Two notes on this table:

1) Joe Cool's +50 is actually a +60 — with his guessing pattern, there's no way on the May 2000 SAT for him to get a combined score of 1510, so I fudged a little.

2) Above +100, the odds drop precipitously, which is why I didn't bother to list them; the odds of the first three examinees doing 150 points better than their “true” score are less than 1 in 50,000. Incidentally, Joe Cool, for whom an additional 140 points means a perfect score, has about a 1 in 2,500 chance of achieving this.

V.

Do those numbers convince you that luck isn't much of a factor on the SAT? Well, if they do, you're a sucker, and you've got to learn to not trust me so much. The table on the last page, though informative, isn't particularly relevant to the most reasonable formulation of the basic argument: students who take the test *multiple* times and are judged only by their highest scores — their highest scores on *each separate section*, which can be mixed and matched to the student's best advantage — are likely to inflate their scores through luck. Is this true?

Before I answer this, I need to dispose with one last piece of hyperbole. Vigdor raises the spectre of a spiraling “SAT-taking ‘arms race’” by citing “a small number [of students who] took the test five or more times.” The problem here is that he's only talking about college *application numbers*; if he examined the acceptance rates, he might find a more pleasing story. The truth, in my experience, is that any student who takes the SAT more than three times isn't screwing over the disadvantaged, they're screwing over themselves. I've worked for two different test prep companies, and the advice for SAT-takers was the same at both: Twice is fine, and is indeed strongly recommended unless you ace the test the first time around. Three times is okay, but looks a little iffy; it should only be done if you've really improved your SAT skills, or if both of your previous tries were genuine fiascos. But four or more times just makes you look like a goddamn weenie. Though any ranking formula will only 'see' the student's highest scores, the admissions officers themselves see the entire list — and they do, in fact, penalize the goddamn weenies. Emily Wolper, who is now an independent admissions consultant, worked as an admissions officer at Columbia for several years, and in an e-mail, she described the conventional admissions wisdom as follows:

There are a couple of concerns regarding an applicant who has taken more than 3 administrations of the SAT I. The first, and most pragmatic, is that applicants usually level off in improvement after taking the test two or three times - it is rare to see a marked increase in score after that point. Also, why would an applicant who has interesting activities and passions to which he/she dedicates significant energy have time to take more than two administrations, let alone four or five? While most admissions offices will take the highest combined SAT I score, there is still doubt cast upon an applicant who has made a career of taking the SAT I.

This position is corroborated by former Dartmouth admissions officer Michele Hernandez's book *A Is For Admissions*:

There is a certain resentment or even suspicion regarding students who take the test many times. There is really no need to take the test more than two or three times at most, in my judgment.

In short, we're not talking about the difference between taking the SAT once and taking it an ever-spiraling number of times, we're talking about the difference between taking it once and taking it two or three times. For that matter, it's barely even *feasible* for a student to take the test “five or more times,” at least not in the year — from January of Junior year to December of Senior year — that really matters. There's seven test dates per year, and any test-obsessed student is going to want to take the SAT IIs at least three times as well.

VI.

Now, back to the original question. Below, I've calculated the odds that each of our hypothetical students, after taking the test three times and picking their best Math and Verbal scores, will find their combined score significantly above their "true" score:

	+ 50	+100
<i>Shermy (1040)</i>	22.7%	0.5%
<i>Linus (1180)</i>	41%	2%
<i>Marcie (1320)</i>	25%	2.3%
<i>Joe Cool* (1460)</i>	42%	2.6%

*With Joe Cool, I've again fudged a little because of the lack of odd scores at the top of the scale: this time, the +50 is actually a +40. Incidentally, Joe has about a 1 in 500 chance of getting 800s in both sections, while everyone else has less than a 1 in 10,000 chance of going up 150 points or more.

So...what do all these numbers, which took me way too long to figure out, mean? I'm not exactly sure. I'm admittedly a little surprised at how high the left-hand column turned out to be, and I'm especially surprised at Linus's +50 being so much likelier than Marcie's — I triple-checked the numbers, but I still can't really explain why it happens. But I think what the table makes clear is that, although small random fluctuations are common — which I assume is one of the reasons ETS claims a margin of error on the test of between 60 to 70 points — larger ones quickly become rare.

Is Vigdor right to consider this a significant problem? Possibly. Admissions officers have the unfortunate tendency to regard even differences within the margin of error as significant — ironically, moreso at the highest levels, where they're most likely to be generated by luck — and this does give an unfair advantage to students who can cherry-pick their best scores from a set of three tests. But is this a system "alarmingly similar to the fictional dart-throwing academy," which provides "advantages to the wealthiest dart players *rather than* (italics mine) the most accurate ones"? No — if anything, it provides advantages to the most accurate players who *also* happen to be wealthy. Though Vigdor's analogy is ultimately silly, it does contain a grain of truth: the metaphorical throw of one dart can indeed mean the difference for the Joe Cools of the world between a 1560 and a perfect 1600. But to get to that point, you've got to nail an awful, awful lot of bullseyes in a row. And if you've done that, there's something more than dumb luck at work.

VII.

Luck *isn't* the only reason that students who retake the SAT tend to get higher scores, though, and the other factors Vigdor brings up are more significant. It's true that students who take the test for a second or third time tend to get higher scores due to, as Vigdor states, "greater familiarity with the test, improved knowledge or ability, and the fact that people who decide to retake the test really were more likely to have had a bad outcome the first time around." Actually, these all strike me as fairly *compelling* reasons to accept the higher score as the more valid one, but on this it appears we will have to agree to differ.

In addition to all this, there is the vast gray area of fate. The examinee may happen to know more (or less) of the vocabulary being tested than they would on average – although there are enough vocab words on each test that, as before, it's unlikely this will cause them stray too far from their "true" score, it could mean another 20 or 30 points. Although the math questions test virtually the same concepts from test to test, the hardest ones also tend to be the oddest, which means that whether a high-scoring student gets the toughest questions in each section will vary somewhat from test to test – again, probably not by more than 20 or 30 points, but it all adds up. Then, there's the factors beyond the control of either the test or the student – screw-ups by proctors, over- and under-heated rooms, Sonic Youth playing an outdoor concert directly outside the testing site (which once happened to me when I was proctoring, though thankfully it was only a practice test – still, very disconcerting,) the kid behind you having a bad case of the sniffles and/or leg-joggles, etc., etc., ad infinitum.

All of this can affect a student's score, and it's true that students who take the test multiple times are more likely to at least once play what the Scottish might call "their usual game." And ultimately, whether this due to luck, skill, or fate doesn't really matter. In any case, the core argument holds: *Given that disadvantaged students are, as Vigdor finds, significantly less likely to retake the SAT than privileged students are, the current admissions policy effectively screws them out of points.* So long as Vigdor's findings are accurate, which I see no reason to doubt, this is not just likely – it's statistically inevitable. There's no way around it. It's built into the rules of the game.

Setting aside the rhetoric, Vigdor has diagnosed a serious problem accurately, and I'm completely in agreement with his goal of leveling the admissions playing field. But he proposes his solution – admissions departments taking the average of all scores, or, alternatively, looking only at the most recent set of scores – as though the costs were nil and the benefits couldn't be achieved in any other fashion. I think he's wrong on both counts.

VIII.

Vigdor's main proposal, in which admissions departments average all the scores, works in the same way that pushing on wallpaper to get rid of an air bubble works; it fixes things, as long as you don't look around too hard. The most obvious trade-off is that, while the current method sometimes unfairly rewards good luck, his solution will, by the same token, sometimes unfairly punish bad luck. But which is worse is one of those false-positive/false-negative debates that's essentially unresolvable, so I'm willing to make a deal: if Vigdor agrees to stop concocting stories about rich dopes that blindly stumble onto an extra 300 points or so, I won't concoct any stories about kids who manage to, through sheer bad luck at guessing, screw themselves out of 300 points or so. It's just as unlikely, and just as unfair. And that's the last time I'll mention luck.

This is, in any case, a side issue. The real air bubble here is the incentives: while Vigdor's proposal decreases the incentive to retake the SAT, which is arguably good, it vastly increases the incentive (right now around 0) to cancel your scores, which is almost certainly bad. Vigdor never brings up canceled scores in his essay — a striking omission, considering that the increase is an inevitable consequence of his proposal, and that it could make a bad situation even worse.

IX.

For one thing, cancellations are simply unpleasant; they increase the net total of unhappiness in the world. Deciding whether or not to cancel is incredibly stressful, perhaps more so than the SAT itself. A student has only three days to decide whether or not to cancel their score, and must do so blindly. Under Vigdor's averaging proposal, if the examinee does poorly and doesn't cancel their score, they've screwed themselves over irrevocably: they'll need to improve their score by exactly twice as much to achieve the same result. If they do cancel, they'll never know whether they threw away a golden ticket, as the score is expunged from their record. I'm no expert on teenage psychology, but this does not strike me as a recipe for improved mental health.

A more fundamental drawback, though, is that the decision to cancel rewards a skill that's pure gamesmanship – the skill of analyzing your own SAT performance and determining your probable score. It's eternally arguable whether or not the skills tested on the SAT itself are relevant ones, but I don't think anyone can seriously argue that a student's Poker intuition should be, and this is what averaging scores turns the process into. Students who know when to hold 'em and know when to fold 'em are the ones who benefit from increased cancellations; students who aren't able to analyze their hand and calculate the odds are the ones who get screwed. And this kind of meta-SAT skill is at least as biased towards the privileged as the SAT itself. To possess it, after all, it helps to have any or all of the following:

- 1) A better understanding of statistics and probability than you need for any actual SAT question.
- 2) A certain cold-blooded detachment from your own performance, including a willingness to have spent 25 bucks and three hours for nothing.
- 3) A decent sampling group of your own performance, which requires having taken – at a bare minimum – a half-dozen practice tests beforehand.
- 4) As much knowledge about the test you just took as possible. Which section was the experimental one – and thus doesn't count towards your score – is the single most important question, but knowing whether or not the test was considered difficult (and thus will have a more forgiving scale) or not can affect your self-evaluation by up to 50 or 60 points.
- 5) A tutor and/or teacher that can analyze your performance and point out patterns and tendencies that you, as a high-school student, are not going to see.

#1 is class-biased, though for no more reason than any math skill is; privileged students, by definition, tend to have privileged educations. #2 is, aside from the dependent clause, too psychologically complex to pin down demographically, but keep it in mind when I talk about "stereotype threat" on the next page. And let it also be said that, for better and for worse, The Princeton Review's guiding ideology – "The SAT is a scam," to quote John Katzman – intentionally fosters this kind of detachment in its customers. #3 is clearly biased in favor of those who either take prep courses or who take lots of practice tests on their own. #4 rewards students who know to go to The Princeton Review's website, on which the company posts an evaluation of each SAT within 24 hours of it being administered, including which section was the experimental. It's to the company's great credit that they don't restrict access to these reports to their customers, but even knowing this exists has to correlate pretty well with socioeconomic privilege. (Incidentally, this service doesn't cover the Sunday tests, leading to a canceling bias against Jews, Seventh-Day Adventists, and anyone else who doesn't take the test on Saturday.) And #5 clearly goes to the highest bidder.

X.

Vigdor approvingly cites the LSAT as an example of his proposal in effect; law schools take the average, he notes, not the highest score, and "the rate of LSAT-retaking is correspondingly much lower than the rate of SAT-retaking nationwide." Putting aside the fact that a group of 21-40 year olds actively seeking a high-pressure graduate program isn't necessarily comparable to a group of 17-year olds who just want to go to college, it's strange that Vigdor doesn't cite any improvements other than the decrease in retakes itself. Is there any evidence that LSAT-takers *prefer* this policy over the SAT's, or that the LSAT's policy has actually *helped* disadvantaged students? I've tutored LSAT students of all economic classes, paying anywhere from \$0 to \$200 an hour for my services, and virtually every one has groused about the averaging policy; even students who don't think they'll retake the test would prefer to walk into the testing site with the knowledge that a lousy performance can't permanently harm them. And there's evidence to suggest black students might be disproportionately harmed by increased test pressure — meaning that, while a decrease in retakes might decrease the racial gap in LSAT scores, the accompanying increase in test pressure might actually wind up *increasing* it.

This isn't an issue of gamesmanship per se; it has to do with students' perceptions of the test itself. Claude Steele, a professor of social psychology at Stanford, has developed the theory of "stereotype threat," which I find the most promising explanation of the racial gap in standardized testing scores as yet. In a 1999 article for the Atlantic Monthly, Steele explains his theory — you should really go read the entire thing, but for the attention-deficient, here's the gist of his argument. "Stereotype Threat" is the well-documented tendency people have to clam/freeze-up/choke when put in a situation where they feel they're being judged according to a negative stereotype. It's a self-fulfilling prophecy, as neurotic as it is ubiquitous; it applies to all races, sexes, classes, etc., so long as they're being placed in a situation where the additional pressure and distraction of possibly confirming a negative stereotype comes into play. The more they care about not confirming the stereotype, the more pressure they'll feel. The more pressure they feel, the more likely they are to underperform — and if underperforming is part of the stereotype, the more likely they are to confirm it.

In the field of standardized testing, this factor is primarily going to work against black and Hispanic students. Whether or not the SAT is actually biased against black and Hispanic students or not is an open question; whether or not many black and Hispanic students *think* it's biased against them is not. And they know that their score will be used to judge them; along with their GPA, it forms the core of their statistical identity. It's true that different people will respond in different ways; some students might overcome the pressure, and a few might even respond positively to it. But on a statistical level, any racial group that, for whatever reason, does relatively poorly on the SAT is trapped in a vicious circle; the statistics confirm the stereotype, and the stereotype itself helps to depress the statistics.

To what extent does this happen? Here's Steele, explaining one of his standardized testing experiments in an interview with PBS Frontline:

What you saw this afternoon is people taking a particularly frustrating test. It's a very difficult test taken from a section of the Graduate Record Examination in literature. We know it's going to cause frustration and that is going to trigger the relevance of the stereotype. When they experience that frustration, they'll sense, oh boy, I could be seen stereotypically here. I could be confirming the stereotype. And for the students you saw who are very strong students, very committed to succeeding in school, that prospect of being seen stereotypically is disturbing. And it can undermine their performance right there. And that's generally what happens. Compared to white students in that situation, they in that situation are not subject to that kind of a stereotype. And so they may be haunted by all kinds of things with regard to performing on standardized tests, but

they're not haunted by the prospect of confirming this stereotype. So you get two groups of students, white, black, who are equally prepared. Equal skills, everything. You give them this very difficult test that is presented as diagnostic of ability. The black student has this extra pressure on performance. And that is in our research invariably reflected in lower performance.

Then you shift conditions just with the touch of a change of the instructions, you present the same test as a test that is something we use to study problem solving in the laboratory and is not diagnostic of ability. That turns the stereotype off for the black student. Now as the black student experiences frustration on this test, it has nothing to do with the prospect of confirming a stereotype or being seen from the standpoint of the stereotype. And if that pressure of being seen stereotypically is enough to depress their performance, then taking off that pressure should increase their performance. And that's what happens in this research. Presenting the same test as non-diagnostic of ability, black students perform just as well as equally prepared white students in that situation.

How big of an effect is this thing having?

A 12 would be the score that white students or black students who are not under stereotype threat get. This is a 25-item test; they get 12 correct. Under stereotype threat, they get about 7 or 8 correct. So that's a very substantial difference. And that's just in a 25-minute section of the Graduate Record Examination. If you played that out over the six sections that usually comprise that exam, that could cause a substantial difference in score.

In fact, if those statistics held as a percentage over the entire SAT, it would mean the difference between a 1030 and an 870, a 160-point gap. Steele's research is fascinating and of vital importance, and I hope to discuss it at other point in much more detail. Its relevance here, though, has to do specifically with that perceptual shift; for all we know, it may very well matter less whether minority students do retake the SAT than whether they think they can. Or, to put it another way: Steele's experiment, in which he told students that the standardized test they were taking – a test of no real-world consequence – was diagnostic of their ability, lowered the black students' scores by the equivalent of 160 points and the white students' scores not at all. What happens when you tell these students that their score, if they choose to know what it is, is not just diagnostic of their ability, but *irrevocable*?

XI.

Vigdor's alternate proposal – to take the most recent set of scores, rather than the average or the peak – is actually better than his main proposal, but it still falls short. While it reduces the "permanent record" anxiety to some extent – and also shifts its burden more to those taking the test for the second or third time – it still increases it from where it is now. And canceling, though not quite as large a factor as with his main proposal, would still be much more important than it is currently.

It's always worth trying to come up with a better admissions system, and I can even imagine a hybrid of Vigdor's alternate proposal and the current state of affairs that might be a genuine, if modest, improvement. For instance, perhaps students who take the test twice (in their Junior or Senior year) get to keep their best scores, but students who take the test three or more times are stuck with whatever they scored most recently. Though a bit clunky, this solution avoids the pitfalls of each extreme: while it discourages students from taking the test a third time unless something went truly awry in the first two, it doesn't so strongly encourage cancellations or conjure up visions of Permanent Records.

But even this is really just another squish of the metaphorical air bubble; at best, it spreads the pain around a little more evenly. Like both of Vigdor's proposals, it's ultimately focused on the wrong end of the argument. Let's take another look at my summary of Vigdor's core thesis: *Given that disadvantaged students are significantly less likely to retake the SAT than privileged students are, the current admissions policy effectively screws them out of points.*

Assuming this is true, which it is, there's only two logical ways to fix it: you can change the current admissions policy, or you can eliminate the given. If the last few pages have demonstrated anything, I hope it's that even the most well-intentioned change in policy will not only have unintended side-effects, but could even completely backfire. If the next few pages are to demonstrate anything, I hope it's that eliminating that odious given is considerably more feasible — and important — than Vigdor seems to assume.

XII.

Why do disadvantaged students retake the SAT less frequently? Contrary to the dart game run by greedy carnies that Vigdor envisions, it's clear from looking at the College Board's website that few if any students are getting intentionally shafted on the basis of cash. Students who use fee waivers aren't second-class examinees; the examinee receives a detailed score report and test booklet that costs the paying students an extra 10 bucks, the waiver can be used towards a late or standby registration (though the student has to pay the actual late fees, which seems fair enough,) the waivers can be used to take the test for a second time (though not a third time, which is unfair,) and applying for the waiver automatically registers the student for application fee waivers at most selective colleges. Having not gone through the process myself, I don't know the whole story: it's possible that the income bar for waivers is set unfairly high, or that the application itself is hopelessly byzantine. But as long as neither is the case, this is a pretty good deal if students take advantage of it.

The problem, as Vigdor points out, is that they're generally not. "SAT fee waivers," he writes, in the only paragraph in which he mentions them, "which can reduce the financial obstacles to retaking the test, fall short of creating a truly level playing field." Why not? Vigdor's doesn't mention the charge for re-retaking, which is indeed probably a minor factor. According to him, the issue is essentially one of opportunity costs: "Fee waivers," he states, "do not compensate for the free time or wages that applicants forgo when they retake the test."

This is undeniable, and a fair point to bring up; if anything, Vigdor drastically undersells his own argument. At the end of the article, he states that "absent strong incentives to retake the test...teenagers would find themselves with one or two extra Saturday mornings." One or two? Any student who seriously wants to improve the second time around will likely need to either take a prep course or do the equivalent amount of work on their own. At the Princeton Review, a course — which runs a cool 1000 dollars, incidentally, with no financial aid information to be found on the website— is 35 hours of classes, 14 more hours of practice tests, and at least 10-15 hours of homework. Discouraging students from retaking the tests doesn't just give them a head start on the weekend; stretched out over a series of consecutive Saturday mornings, all of these hours would span almost half the year.

But opportunity costs cut both ways. The same argument Vigdor makes against the SAT can be made against teachers who allow students to rewrite papers and take the higher grade, or to retake tests they choked on. The same argument, in fact, can be made against *any* factor in which additional time spent tends to improve performance, which includes everything from application essays to extracurricular activities to GPA itself. And if you start looking at GPA with the same critical eye for opportunity costs that Vigdor brings to the SAT, it quickly becomes unclear whether disadvantaged students are retaking the SAT less frequently because it's not worth their time, or because nobody's bothered to mention to them that, in sheer bang-for-your-minute terms, it's the best deal they've got.

XIII.

Imagine, for the moment, that you're a disadvantaged but talented young athlete who has just finished his or her Junior year of high school. It's the first day of summer vacation, and you received two letters in the mail: one containing your final grades and cumulative GPA, the other containing your score on the May SAT. Both are looking pretty weak; your cumulative GPA is a 2.0, and your SAT scores are a combined 820. By a remarkable coincidence, this puts you at the bare minimum of both ends of the Division I eligibility scale we first saw on page 2, which means you're not eligible at all: with a 2.0 GPA, you'll need a minimum SAT score of 1010, and with an 820 SAT, you'll need a minimum GPA of 2.5. Something has to be done, or else you won't be going to college. Let's now assume, for purpose of direct comparison, that the only options open to you are at the extremes: either you retake the SATs and try to improve by 190 points while maintaining your 2.0 GPA, or you keep your SAT score and try to raise your GPA to a 2.5. Let's also assume, for the sake of not opening up an endless can of worms, that which option would most improve your character, teach you valuable life skills, make your parents happy, encourage positive political change, please your chosen deity/-ies, etc., etc., is of absolutely no concern of yours; your only interest is having as much time for non-GPA/non-SAT related activities as possible. Which option should you choose?

To be honest, it's not even a contest — any economically rational individual who is shown the numbers is going to immediately head for their nearest library's test prep section. Ironically, your saving grace — the only thing that gives you any chance at all — is that Vigdor's proposal doesn't apply to you: even if admissions departments were convinced to average their SAT scores, any minimum eligibility requirement is still only going to look at the highest SAT score. This isn't to say that going from a 820 to a 1010 on the SAT isn't a huge undertaking; it most certainly is. Only 25% of students who enroll in a Princeton Review course go up by this much, and you, of course, don't have the money for a course, so you're going to have to eat up a whole lot of time studying. To be cruel, let's say it'll take you 120 hours — about twice as long as a Princeton Review course plus its attendant homework — which comes out to about 2 hours of studying, 5 days a week, for your entire summer.

This sounds pretty rough, and it is — but compared to making the equivalent improvement in your GPA, it's a cakewalk. By the time Senior year rolls around, turning a cumulative 2.0 GPA into a cumulative 2.5 is like steering a plummeting cow into a wading pool; it can be done, but you're not leaving yourself much room for error. In fact, considering you'll need a perfect 4.0 for your entire Senior year to accomplish this, you're not leaving yourself any room for error at all...nor, one would presume, with a minute of free time for your entire Senior year. The optimal solution, of course, is the Epicurean one: improve your SAT score by 100 points, maintain a 3.0 GPA for your Senior year, and you'll meet the eligibility requirement in less time while learning more in the process. But if you had to choose between door #1 and door #2 — or, more precisely, between door #4.0 and door #1010 — it's pretty clear which will give you a fighting chance.

XIV.

It takes about 10 minutes to make the case to a student from a low-income family that preparing for the SAT and taking the test two or even three times is in their own economic self-interest. I know this, because I've taught a number of pro bono SAT classes to 2nd semester Juniors over the last decade, and that's about as long as I need to spend at the beginning of the first session before the students all get the picture of what I expect from them. The example I use is a somewhat sunnier variant on the previous page; instead of discussing minimum eligibility requirements, I bring up the academic index used, as Michele Hernandez describes in *A Is For Admission*, at all of the Ivy League Schools. This score consists of three equally weighted parts: SAT scores, SAT II scores, and GPA. Tossing the SAT IIs away as next semester's problem, this leaves a student with a score that, as on the previous page, is half SAT and half GPA. To get the weighted value of the GPA relative to the SAT, you can multiply the GPA by 400. (It's actually a bit more complicated than this, as class ranking can nudge the GPA value in either direction, but let's assume that's not a factor). Thus, a 4.0 is equivalent to a perfect 1600, a 2.0 is equivalent to an 800, a 2.5 to a 1000, and so on — numbers pretty close to those used in the Division I chart.

As my hypothetical, I present a student who's moderately strong in both areas: a 3.0 GPA and a 1200 combined SAT. And then I just do out the math on the blackboard. That 3.0 GPA is cumulative over five semesters; since 2nd semester senior grades aren't really used in college admissions decisions, this gives the student two more semesters to pull their GPA up. In order to go up the equivalent of 50 points on the SAT, they'll need to average a 3.4 for the next year. In order to go up 100 points, they'll need to average a 3.9.

Once it's pointed out that going from a 1200 to a 1320 is better for one's academic index than getting straight A's for the entire next year, the discussion generally turns to more practical matters, like how one goes about improving their SAT score by 120 points. But sometimes they argue, and many of their arguments are convincing. They can convincingly argue about whether the Academic Index is fairly weighted — I mean, are these three factors really exactly as important as one another? — and my students often do. They can also somewhat convincingly argue that the formula should weigh the final two years of GPA more heavily than the first two years, and my students often do. But there's no way to convincingly argue, as Vigdor attempts to, that for a time-strapped student the SAT isn't the more economically rational deal, and my students never do.

XV.

Of course, people are rarely economically rational, and high school Juniors doubly so. This isn't a dis; I'm a 28-year old college graduate, and between the can of caffeinated corn-syrup water sitting next to my keyboard and the (possibly literally) deafening music coming through my speakers, my behavior has to be at least 66 2/3% economically irrational as I write this. Even when given perfect information, people will often pick the wrong option, and there's no real way to fix this; it's called free will, or a reasonable practical facsimile thereof.

But I suspect that, at the moment, students are receiving nowhere *near* perfect information, and this is simply unacceptable. What percentage of educators know about and take steps to ameliorate stereotype threat? What percentage of school counselors — whom students and parents are effectively required go through in order to apply for fee waivers — actively encourage qualified families to take advantage of them? What percentage of counselors inform them that retaking the test is probably worth their effort? What percentage of guidance counselors even make sure to actively remind qualified families that SAT fee waivers exist? I have no idea. But unless efforts are made to level the playing field, it's safe to assume that the worse the school, the lower the percentage — a factor that could impact retesting rates just as much as any SAT admissions policy. The problem with current SAT policy might not be, as Vigdor argues, that it's drastically unfair. It might be — as Steele has shown could be the case with the test itself — that nobody has bothered to inform anyone that it isn't.

This theory, I'm sure, will initially strike many as loopy, cuckoo, or perhaps even 'bananers'. So let me illustrate my point by talking about an situation that I think we can all agree is eminently fair: need-blind admissions.

"Need-blind admission," first adopted by Yale in the late 60s, is a purely egalitarian policy and one of the crowning moral achievements of our educational system. What it means is that there's a firewall between the admissions department and the financial aid department; at need-blind schools, students are admitted and rejected by officials who have no idea what that student's financial status is. I'm proud to say that Amherst College has need-blind admissions, both because I was there for the protest when the administration tried to get rid of it in 1992, and because, having been on over \$18,000 a year of financial aid myself, I might not have gotten in there otherwise. I'm sure that, as a member of its faculty, Vigdor is also proud of Duke's need-blind admissions policy, as well he should be. Not having a need-blind admissions department is simply wrong, and if you don't believe that, I'm shocked you've bothered to read this far. I can imagine exceptional circumstances in which a cash-starved school might need to temporarily let down the firewall between admissions and finance, but for well-endowed schools such as NYU and Mount Holyoke to lack need-blind admissions departments is shameful on principle. The problem, however, is that from a practical standpoint, the moral high ground is not what matters. What matters is whether need-blind admissions matters at all.

XVI.

The odd truth, according to Robert Shireman's article "Need-blind Admissions Policies: How Much Do They Affect The Enrollment Of Lower-Income Students?", is that when you look at the percentage of students on Pell Grants at the most competitive colleges — a reasonable enough way to determine a need for financial aid, I assume — Mount Holyoke and NYU rank near the very top. Mount Holyoke has 23.8% of its students on Pell Grants, highest of any liberal arts college; NYU has 18.4%, which, except for the four Universities of California, is tops among national universities. Overall, highly competitive need-blind schools actually enroll Pell recipients at a *lower* rate than do highly competitive non-need-blind schools. Even factoring out the curve-breakingly Pell-heavy U. Cal. system, non-need-blind national universities have 12.4% of their student body on Pell Grants; for need-blind schools, it's less than 11%. At liberal arts schools, the numbers are around 13.6% for non-need-blind schools, and 12.5% for the need-blind. Duke, despite its need-blind status, is near the bottom of the list at 9.4%. Something missing there, just as something is missing in the SAT process.

In both cases, I suspect, it's some form of active outreach. That's certainly what Shireman believes; every single one of his proposed strategies for increasing low-income enrollment has this at its core. His #1 strategy is simply "Information":

Lower-income families need to hear, again and again and again, that they don't need to be rich to go to a high-tuition college or university. And they need to see the elite institutions not only as desirable places to go, but as appropriate places for people like them. As researcher Patricia McDonough notes, "students develop college plans based on their families' and communities' values and assessments of appropriate goals." It can take a lot for a child whose parents never attended college to feel that Stanford or Yale is a viable option. For someone who is not one of the elite in society, the concept of going to a prestigious college often just doesn't compute.

This is exactly what needs to be done with the SAT as well, no matter what the retake policy is. It's the responsibility of high school counselors to make students aware of the advantages of using all their fee waivers, it's the responsibility of college admissions offices to encourage these students to take the SAT again if it's going to help them, it's the responsibility of ETS to increase the number of waivers available to each student for as long as there is incentive for students to take more than two SATs (and they should give a voucher for a copy of 10 Real SATs with every fee waiver, too) and it's the responsibility of entities like this website and The Princeton Review to make as much information available for as little cost as possible. Not doing this is tantamount to suppressing scores. Even if the test itself is a level playing field, an information imbalance will inevitably tilt it, just as even if an admissions policy is itself fair, it doesn't necessarily lead to fair results:

Financial aid is a necessity, and the Need-Blind Admissions-style commitment to financial aid is certainly commendable, but the policy alone is insufficient; by itself, it will not bring economic diversity to elite universities. And there is a danger that an excessive focus on NBA as the critical policy can cause those institutions to feel like they are off the hook: "If working-class kids don't apply, don't meet our standards, or don't come, it's their own fault, since the financial aid offer is so generous." Instead of praising institutions for adopting need-blind policies, we should praise institutions for results: the enrollment of lower-income students.

The reasons for this benign neglect are only possible to discern through telepathy. A Panglossian might say that this is because colleges are structurally good at setting internal policy

and bad at doing external follow-up, while a more cynical person might compare the policy to President Bush's clever method last year of cutting Veterans' health benefits without actually cutting them: just stop telling the Veterans that the benefits exist. Either way, Shireman is right; what counts isn't policies, but results. And in the case of the SAT, it's the results, not the policies, that are most in need of fixing.

XVII.

Let's recap. Luck isn't a huge factor on the SAT, although taking the test two or three times does give students an advantage for other reasons. Vigdor's shown that disadvantaged applications do tend to have a significantly lower retake than privileged students, effectively suppressing their scores, while I've shown that Vigdor's proposal for fixing this has a significant downside he didn't mention. Steele's theory provides evidence that standardized testing in and of itself may not be racially biased, but that perceptions of it do create reality — that it can be the frame, not the game, that's the problem. I've shown that, from a sheer cost-benefit analysis point of view, as long as the SAT continues to be a major factor in college admissions, it's to virtually any student's benefit to spend time preparing for it. And Shireman has shown that the realm of external outreach, rather than that of internal policy, can be the deciding factor in whether a college attracts a diverse student body.

So what's the disagreement? My proposal and Vigdor's are like the famous vase-face optical illusion; we both agree on the same basic goals, we both agree on the same basic data, and we both come to basically opposite conclusions. There's a simple reason for this; Vigdor thinks the SAT is an bad test, while I think it's a good one. He makes his opinion explicit near the end of the article, when he states the only benefit of his proposal that I find completely plausible: that discouraging retaking will leave American high-school students with millions of hours of additional free time each year to, among the other more ambitious activities listed, "just sleep in." This is, as I said, completely plausible. As long as you think students are spending millions of hours a year on an activity that's below some baseline level of worth, then by definition they will benefit from not having to do it anymore — and Lord knows that if there's one thing that high school students lack in this day and age, it's millions of hours of free time. It's not as though I don't sympathize with the argument; if I were to look at all the homework I did in high school, I imagine I would make the same one for at *least* 30% of it. But I simply don't agree with it here. This is Vigdor's core assumption — if you think the SAT is an inherently bad test, then of course you're going to view any policy that leads to an increase in retakes as, as he puts it, 'an arms race' — and the problem is that he never gives the reader any reason to believe this is the case. I understand why, as it's for the same reason I'm not about to argue the other side: to explain why I believe the SAT and (some) other standardized tests have (some) value simply can't be explained at the end of an article, and I'm not even positive that I'm right. But his entire proposal rests on this as an assumption, and he needs to make it overt in some way other than slipping in little zingers.

More importantly, if he's going to keep making arguments about the SAT, he also needs to keep his assumptions straight. The truly strange thing about Vigdor's initial analogy, so carefully contrived to cast the test in doubt, is that it's based on the opposite assumption as his own proposal is. The problem with the Dart U. exam isn't that it's broken; as I've shown, Vigdor's set-up could be vastly improved simply by making it actually resemble the SAT. But this superficial issue just distracts from the analogy's core flaw: the dart-throwing exam is *too good*, far

better than even the staunchest pro-SAT advocate would possibly argue. After all, it's pretty easy to define what an Academy of Darts is looking for: quantifiable skill at throwing darts. At least in theory, there *could be* a perfect entrance exam for such a school, because there's no argument about the goal itself — all you'd have to do is hash out the game theory and you'd eventually come upon the most equitable method. But no sane person believes this about the SAT, because there's simply no corresponding SAT Academy; until colleges decide to turn all their classes into non-stop bubble-a-thons — or until ETS adds a section in which the examinees write a 7-page paper entirely through using Google from the hours of 3 to 8 am while half-drunk and jacked up on Ritalin — there will always be a mismatching of skills, and thus a dissonance.

And because the SATs are rightly and/or wrongly perceived as being about more than just predicting freshman GPA — because they're so woven into the national debate about education — this dissonance isn't fundamentally resolvable, because it's not based on the rules of the game, it's based on what we value as a society. The big question, in other words, is not whether we can determine if someone's good at darts; it's whether *being good at darts is a good thing to be good at*. Is it? As a person whose values are both leftist and (relatively) pro-testing, I know I'm in the minority on this issue, and I look forward to joining that debate through this website. I also look forward to seeing Vigdor's full report; as adversarial as this essay may seem, I fully credit him for bringing an issue to my attention that I had never stopped to consider before. But I hope that in the report he lays out his assumptions more explicitly — and that if he decides to start off with an analogy, it's one that's a little more on target.