

*The Fairness Hoax:
Why the case against the SAT defies logic.*

Patricia Hausman, Ph.D.

presented at
The SAT: Its Nature and Role
Virginia Association of Scholars
Charlottesville, Virginia • October 13, 2001

I want to thank Avent and the board for all their efforts to make this event happen. It gives me another opportunity to be on the unpopular side of a contentious issue. That issue, of course, is the SAT. Actually, its official name is now the SAT I Reasoning Test, but I'll continue to call it the SAT for short. My position on the test is simple. It is really getting a bum rap.

I also believe that respect for the SAT is more the rule than the exception among academics, though they may be reluctant to say so openly. After years of controversy, relatively few four-year colleges and universities have made it optional, and of these, many were not particularly selective in the first place. That the vast majority of schools still require the SAT or a similar test speaks volumes to me.

I wonder how many believe the reason for the SAT's continued use is that a vast, right wing conspiracy controls our universities. Perhaps a few, but I doubt very many. A far better explanation is that even the Educational Left understands that, to paraphrase Philip Dick, "reality is that which, when you try to stop believing in it, doesn't go away." I could try to believe that the SAT is a uniquely unfair aspect of the admissions process, or that we really don't know what it measures. The problem is reality. It keeps getting in my way.

When I do a reality check, I find no mystery about what the SAT measures. It measures verbal and mathematical reasoning skills. When I confront fairness arguments, reality keeps throwing inconvenient facts my way. I'd like to say a few words about these arguments first.

We do have one undisputed fact in this debate. It is that unless willing to assign seats by lottery, schools with more applicants than places have to set admissions standards. In this context, the charge that the SAT is unfair is not very meaningful. It begs the question, "not fair compared to what?"

Aside from those who propose replacing the SAT with standardized achievement tests, the obvious answer from SAT critics is that the test is unfair as compared to high school grades and factors sometimes described as "holistic criteria." Some of the latter do have a long history of use by admissions offices. Others are much newer, and a few almost unbelievable.

Typically, critics argue for these alternatives on grounds that the SAT produces differences in outcome by demographic group. True enough. But the implication is that grades and holistic criteria do not. This is not true at all.

Grades are a prime example. Low and high socioeconomic groups do not leave high school with identical GPA. Nor do males and females. In California, average GPA in high school varies by as much as 30% among ethnic groups. Differences in curriculum are even more dramatic. Asians are more than twice as likely as underrepresented minorities to take the coursework required for admission to the University of California. Yet critics of the SAT do not cry foul over these differences.

They do complain about the ability of the SAT to predict college performance. They claim it is unimpressive. Again, the obvious implication is that holistic criteria are better. And again, we have a double standard. Many holistic criteria have been embraced with no evidence of their validity at all. Others have some research support, but hardly can claim comparability--let alone superiority--relative to the SAT or high school grades.

Is this logical? Does it make sense that people professing concern about bias despise a computer-scored test, yet embrace fuzzy criteria susceptible to all manner of human bias and even fakery?

Actually, I'm afraid it does. There is a common theme running through these arguments. It is the substitution of subjectivity for objectivity. The less the admissions process depends on objective standards, the more it can be molded to suit the agenda du jour, whatever that happens to be. Admittedly, some call this fairness. Yet I suspect that if given similar facts in a different context, the same people would take a very different position, recognizing objective standards as essential to fairness.

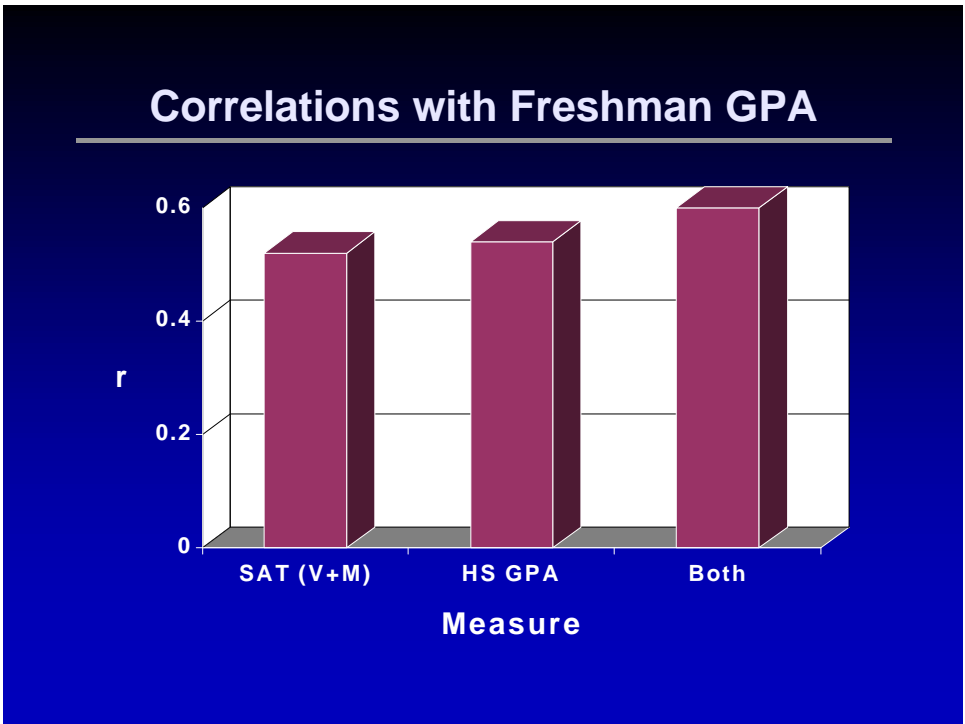
The Nature of the SAT

I assume that all of you are familiar with the basic characteristics of the SAT. What I would like to focus on now is its validity, specifically on what statisticians call predictive validity.

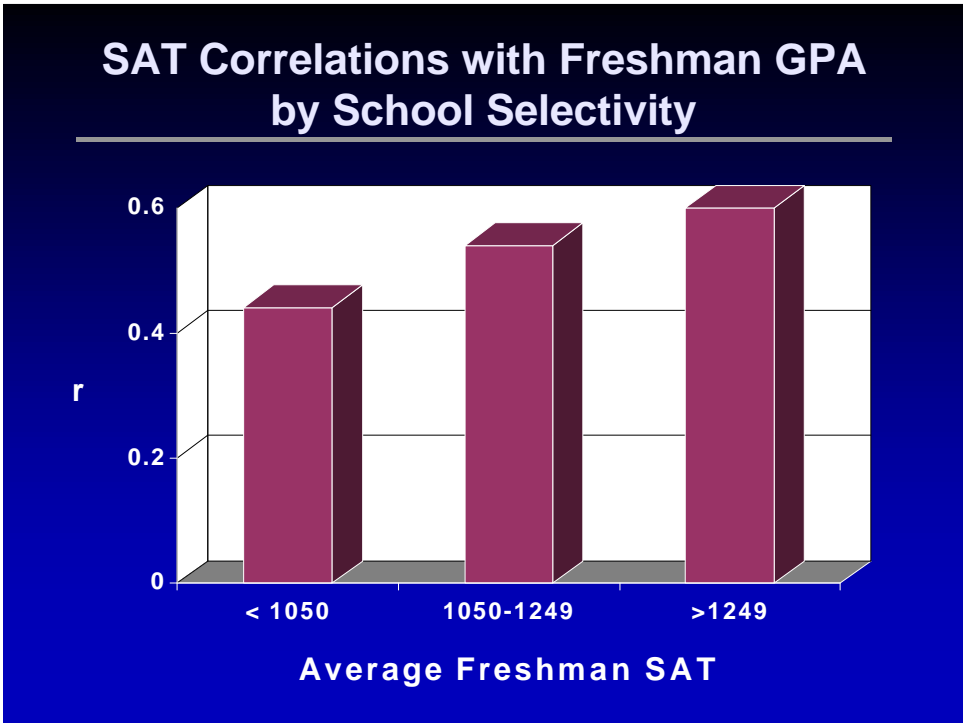
The predictive validity of a test is its correlation with some future outcome. The test is referred to as the predictor, and the outcome as the criterion. In a perfect world, the correlation between two would be 1.0. In reality, correlations of this type rarely exceed 0.6, and often come in far lower. Predictors that have low correlations with outcomes can nonetheless be extremely useful in some circumstances, especially when no better alternative exists. This, too, is a "compared to what" situation. It makes no sense to deride a predictor that has a moderate correlation with future outcomes if all the alternatives have low correlations.

In SAT research, the most commonly studied outcome is freshman grade point average. I used to see this as short sighted, until I learned that there are some compelling reasons to prefer it. Research has also found notable correlations between the SAT and other measures, among them class rank; cumulative GPA; and acceptance to doctoral study, law or medical school. But freshman GPA remains the mostly widely studied outcome.

Recent data show a correlation of .52 between SAT scores and freshman GPA. These figures have been corrected for range restriction, a problem that occurs when a group is pre-selected for the characteristic in question. Restricting the range of talent tends to obscure correlations between predictors and outcomes, but statisticians can make allowances for this effect.



The same study reported a correlation of .54 between high school and freshman GPAs. I think the meaning of a difference this small is debatable, but in some studies it is a little larger. This makes sense to me; a high school record reflects three or four years of observations, the SAT, just three hours worth. What I think most important is that as predictors go, both SAT and high school grades are quite high. Also noteworthy: the correlation between SAT and freshman grades is stronger at more selective schools.



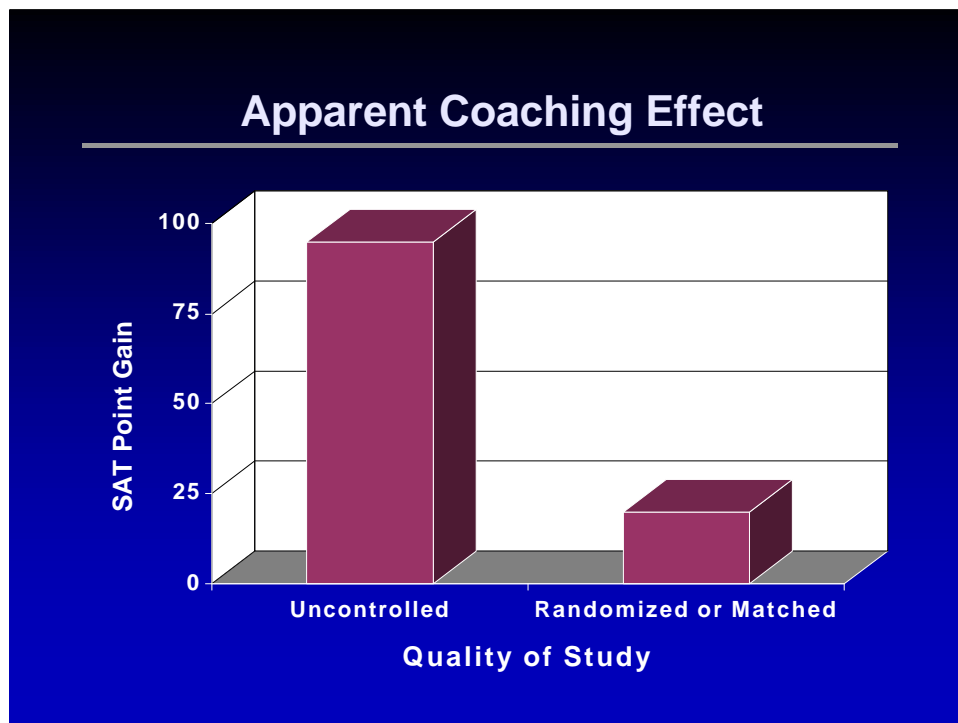
It is only fair to ask why these tools do not bring us closer to perfect prediction. For the SAT, the most obvious explanation is that college grades reflect not just reasoning skills but also qualities that the SAT makes no claim to measure, such as conscientiousness, to name but one.

But other reasons are not obvious at all. As a result, less than perfect prediction is often attributed solely to a test's shortcomings. This is a fallacy. The quality of data used to measure outcomes also influences the predictive validity. With the SAT, anything that compromises the reliability of college grades can make the test appear less valid than it actually is.

Differences in grading standards among departments are a prime example. Students with high scores tend to cluster in more rigorously graded courses, and vice-versa, making the magnitude of the SAT's association with grades more difficult to capture. Another factor is the practice of including courses in performing arts or physical education in the calculation of college GPA. Grades in these courses correlate much less with SAT scores than those in disciplines like math and science. Including them pulls down the overall correlation between the SAT and college GPA.

Validity studies have rarely taken factors like these into account. A few have, and report higher correlations between the SAT and freshman grades than the numbers I showed you earlier. In one such study, the correlation after control for such factors was .65--obviously quite high.

I should add that grades are not the only relevant measure of academic achievement, nor academic outcome the only indicator of educational success. I have focused on GPA here simply because of time constraints.



Before I move on to the role of the SAT, let me address the belief that the test is highly coachable. It does not come from the scientific literature. Research shows that huge

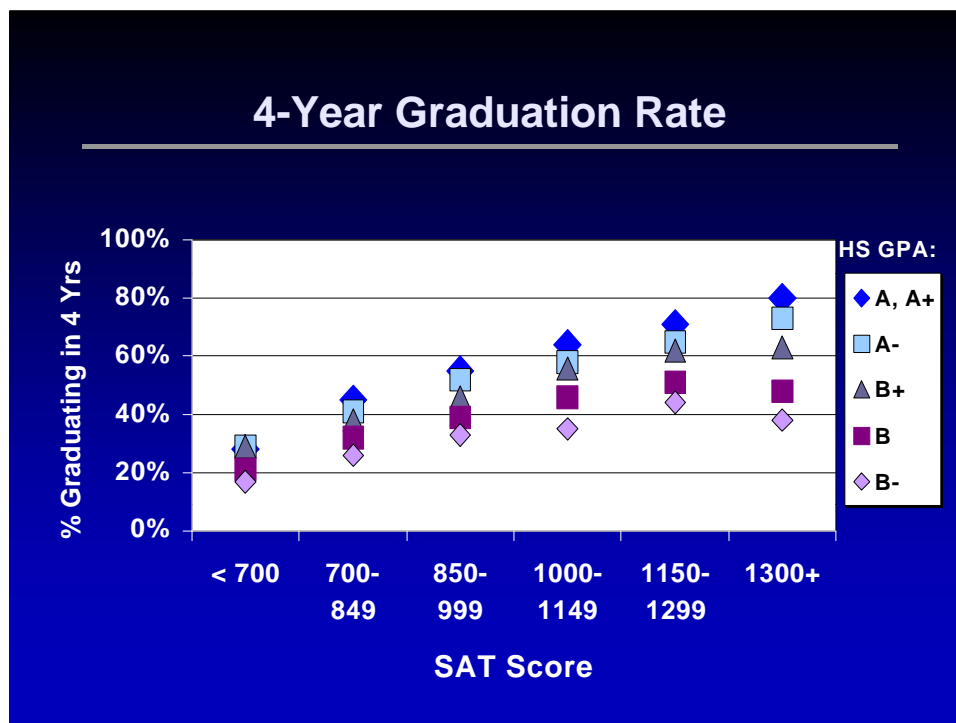
score increases attributed to coaching are more likely the result of poor research design, particularly the absence of a control group. Without one, all gains appear attributable to coaching. In fact, uncoached students often have gains as well. More rigorous studies in which coached and uncoached groups are matched on key characteristics yield very different outcomes. Taken as a whole, controlled studies of coaching show that the effects are real but modest--20 to 30 points, on average.

Again, however, we need to ask the question "compared to what?" If coachability is undesirable, the SAT has an advantage over achievement tests, which research has found to be more coachable. At the same time, I wonder how logical it is to advocate abolishing the SAT on grounds that coaching can affect scores, yet favor tutoring to help students earn better grades.

The Role of the SAT

In one sense, the role of the SAT is like that of any standardized test--to provide a common yardstick for assessing all participants. Grades cannot do this. Standards vary too much from one location to the next.

Grades remain impressive predictors of college success nonetheless, and some ask if the SAT adds enough value to justify the rancor over its use. I think the best way to answer--and appreciate another of its roles--is to consider that more than a third of SAT takers now report A averages. At some colleges, huge percentages of applicants have a GPA of 4.0 or higher. Since all cannot be accepted, distinctions must be made among them, often with limited personnel.



The value of the SAT becomes obvious in this context. Dramatic differences in the likelihood of finishing college in four years have been found among students who enter college with A or A+ averages--from a low of 28% for those with SAT scores below 700 to 80% among those with scores of 1300 or higher. The same general pattern is seen

throughout the top four grade ranges. This kind of evidence shows the enormous value in considering grades and test scores rather than grades alone.

Test scores also help admissions offices evaluate students from schools that do not use traditional grading systems or those that have been home schooled. But in the final analysis, assessment is far more than a process that helps institutions make decisions about individuals. It also allows individuals to make decisions about themselves.

If I have learned one thing from work and personal experience alike, it is that career choice is a profoundly personal decision. I cannot in good conscience advocate that everyone make exclusivity the important consideration in choosing a college. In fact, I did not opt for the most selective school that would have me. Prestige is paramount for some, but as a general rule, the best advice is to pick a college and a career that is a good fit with your personal preferences, your values, your interests, and your abilities. Anything that helps people make this match is information that can serve them well. I see no reason not to put the SAT in this category.

Admittedly, some people go overboard trying to improve their children's SAT scores. But is this an indictment of the test, or a reminder to ask whether we would behave differently under alternate scenarios? What if we replaced the SAT with a different criterion? Would there be any less irrational demand for services that promised to improve it? I doubt it.

What if we were to ban keeping score in sports and all tests like the SAT? Would we really succeed in convincing anyone that humans do not vary in athletic or academic potential? I think this would be about as successful as banning thermometers, then trying to persuade people that it isn't hot in Phoenix. Personal impressions are always an option when scientific measurement is not. The idea that this is good educational policy is lost on me.

In the final analysis, we have to decide whether we are going to shoot this messenger called the SAT. Until we have another messenger that is as accurate, versatile, and cost-effective, I hope we will think very carefully before we fire.