Signaling, SAT Coaching, and Selective College Admissions

Jun Ishii *
Amherst College

Travis Chamberlain
London School of Economics and Political Science

September 2010

Abstract

We investigate how SAT coaching (also known as “SAT Prep”) affects the value of the SAT as an evaluative tool, vis-à-vis other applicant signals such as high school transcript and letters of recommendation. Theoretically, we develop a simple “meritocratic” model of selective college admission that illustrates how SAT Prep may alter the emphasis a college places on SAT scores, depending on perceived SAT Prep efficacy and participation. Empirically, we review existing empirical studies of SAT Prep and conduct our own analysis, using new data from the 2006-07 Orange County (CA) operation of a major SAT Prep firm, to explore actual SAT Prep efficacy and participation. We find that SAT Prep seems to help disproportionately students who are weaker in the tested aptitudes but have academically achieved more, as measured by grades (GPA) and observable quality of attended high school. Our research suggests, perhaps counter-intuitively, that the presence of SAT Prep may provide modest incentives for selective colleges to emphasize the SAT more in their admission deliberation.

KEYWORDS: SAT, College Admissions, Signaling, Hidden Actions

*Corresponding author. 309 Converse Hall, Amherst, MA, 01002 E-mail: jishii@amherst.edu. We thank Ameer Erfan for key data assistance, Steve Trask and Mingzi Shao for excellent research assistance, and Adam Honig, Kate Sims, Marigee Bacolod, and seminar participants at UC-Irvine, Amherst College, Midwest Econometrics Group Meeting, and Yale for useful comments. Part of the paper was written while Travis Chamberlain was an MSc student in Philosophy and Public Policy at LSE
1 Introduction

In recent years, headlines trumpeting low acceptance rates at selective U.S. colleges have become a regular feature of early April newspapers. This year (2010), Harvard College reported an acceptance rate of 6.9%, Stanford University 7.2%, and Massachusetts Institute of Technology 9.7%, all record lows. Discussions in newspapers and other popular media reveal a growing concern about the increasingly competitive college admissions environment, particularly from students and their parents. With falling acceptance rates, students face an incentive to spend even more resources and effort toward improving their college application. One such effort may be participation in SAT Coaching, also known as “SAT Prep.”

The national, standardized “SAT” exam, administered by College Board, is a key signal applicants send to U.S. selective colleges. But the admission process for U.S. selective colleges involve applicants submitting a portfolio of signals indicating academic merit, only one of which is the SAT. This gives U.S. selective colleges who believe that SAT Prep is unfavorably distorting SAT scores the option to de-emphasize the SAT and rely more heavily on other signals, such as high school transcripts and letters of recommendation. We seek to study how perceptions about SAT Prep may affect the relative weight U.S. selective colleges place on the SAT in their admission deliberation. We do so by drawing upon insights from information economics and basic decision theory. We explore these insights further using existing and original empirical analysis on SAT Prep efficacy and participation.

The importance of the SAT to college admissions has led to the formation of a commercial SAT Prep industry, one in which firms and independent tutors coach applicants toward higher SAT scores in return for an often hefty fee. Kaplan, a leading national firm in the industry, reported net revenues of $2.03 billion, $2.33 billion, and $2.64 billion in 2007, 2008, and 2009, respectively. The link between greater selectivity at prestigious colleges and booming private industry in preparatory services for college entrance exams has been documented and studied in other countries, notably those in East Asia. Additionally, the U.S. SAT Prep industry itself has been studied, mostly in terms of its efficacy in raising scores for its clients. In these studies, the test is the only admission factor (non-U.S. colleges) or is analyzed as though it is the determining admission factor. Under such a perspective, the college is limited in its response to test prep; the college may compensate the test score of applicants who are disadvantaged by the availability of test prep (as from limited test prep access) but only to the extent that the college can identify such disadvantaged applicants.

We view SAT Prep as a hidden actions problem that further complicates a college’s admission problem. The college uses the signals in the application to infer an applicant’s uncertain academic merit; applicants are ranked by their perceived merit and accepted in that order. But some applicants may use hidden SAT Prep to distort their SAT signal in their favor. The college, unable to identify which applicants have distorted their signal, may be tempted simply to

---

1See, e.g., http://voices.washingtonpost.com/campus-overload/2010/04/college_acceptance_rates_down.html
2SAT refers to the SAT I exam. NASCAC (2006) reports standardized admission tests as considerably important to the admission decision, second only to grades in college preparatory courses. See Chapter 4
3A similar argument about other SAT biases underlies the current “SAT optional” movement among some colleges
4From the 2009 Annual Report of the Washington Post Company, parent company of Kaplan
5See Bray (2009) for an overview of the international literature. We discuss the U.S. literature later
de-emphasize the SAT. But the college is constrained by the fact that the SAT may signal information not easily gleaned from other signals: the conventional view of SAT as being more informative about “aptitude,” high school transcripts about “achievement,” and letters of recommendation about “character.” By de-emphasizing the SAT, the college may also de-emphasize “aptitude” in applicant evaluation. Furthermore, SAT Prep itself may be correlated with other merit characteristics, altering the informative content of the SAT score for those characteristics as well.

Using a simple model of “meritocratic” college admissions, we demonstrate how a college’s beliefs about SAT Prep may affect the admission weight the college places on the SAT. Specifically, we focus on the college’s beliefs concerning how SAT Prep participation and efficacy vary with the unobservable applicant merit characteristics – aptitude, achievement, character. Perhaps counter-intuitively, SAT Prep may provide colleges an incentive to emphasize the SAT more under certain beliefs. SAT Prep perceived as favoring applicants that are weaker in aptitude but stronger in other merit characteristics (achievement, character) may induce colleges to weight the SAT more in their admission deliberation. Such perceived SAT Prep suggests that SAT score differences understate, in expectation, the underlying aptitude differences and assuages concerns that valuing SAT scores detracts too much from other merits. But for other beliefs, such as SAT Prep efficacy and participation being invariant to applicant merit characteristics, colleges face the more conventional incentive of reducing the admission importance of the SAT.

We review the results from existing empirical studies on SAT Prep, to see how SAT Prep participation and efficacy actually vary with applicant merit characteristics. We supplement this discussion with our own original analysis using recent data from the Orange County (California) operations of a major SAT Prep firm. Our empirical exploration suggests that the average efficacy of SAT Prep is largely invariant with respect to aptitude but greater for higher achieving students, with achievement measured by grades and observable quality of attended high school. Participation, on the other hand, does seem to vary by tested aptitude, with students who otherwise would have tested lower more likely to participate. The data, combined with our theoretical discussion, seem to suggest that SAT Prep may actually provide a modest incentive for selective colleges to emphasize the SAT more in their admission deliberation. We discuss possible future research in our conclusion.

2 A Stylized Model of “Meritocratic” College Admissions

Consider the simple model of one college offering $M$ admission slots to an applicant pool of $N > M$ prospective students. All $N$ applicants satisfy the minimum requirements for admission (i.e. without slot constraints, the college would admit all $N$ applicants) and would accept admission if offered by the college. In this model, the college admission problem simplifies to a rationing problem, with the college struggling to identify the $M$ applicants with the highest academic “merit.”

With only a single college, this model abstracts away competition among colleges and strategic applications by prospective students. This is in contrast to the existing economics literature on college admissions, which builds off the seminal work of Gale & Shapley (1962). Recent versions

---

6We model each applicant’s appeal to the college as being independent of those of other applicants. This abstracts some considerations, like diversity, which make each applicant’s appeal contingent on those of others.
extend Gale & Shapley (1962) by introducing uncertainty; both Nagypál (2004) and Chade, et al (2009) introduce uncertainty in the college evaluation of an applicant. In the models analyzed in such papers, applicants are assumed to submit a single exogenous signal to each college. We propose a different approach, one where we simplify the “market for college education” but enrich the signaling. We assume that applicants provide a portfolio of signals and that the joint distribution governing the realization of these signals may be affected by hidden applicant actions. Our study focuses on the signaling aspect of the college admission problem, rather than the matching.

Suppose that college admissions are based on some set of underlying applicant characteristics which, together, form the applicant’s academic merit. Without much loss of generality, let there be three such characteristics for each applicant \( i \), measured in standardized units: “Aptitude” \( (X_i) \), “Scholastic Achievement” \( (Y_i) \), “Character” \( (Z_i) \). Moreover, the academic “merit” of applicant \( i \) \( (S_i) \), in the eyes of the college, is a weighted sum of these factors:

\[
S_i = \omega_x X_i + \omega_y Y_i + \omega_z Z_i
\]

The linear combination assumes a constant trade-off among the applicant characteristics. We make this assumption for both expositional and algebraic simplicity. The weights \( \omega = \{\omega_x, \omega_y, \omega_z\} \) have the natural interpretation of how much the college values a particular characteristic vis-à-vis the other characteristics. We remain agnostic about \( \omega \) and take them as exogenously fixed, similar to preferences in a standard consumer demand model.

In the presence of complete information, the college would offer admissions to the \( M \) students with the highest \( S_i \) values. But the college does not observe the student characteristics \( \Omega_i \equiv \{X_i, Y_i, Z_i\} \) directly; the college observes \( \hat{\Omega}_i \equiv \{\hat{X}_i, \hat{Y}_i, \hat{Z}_i\} \). Consequently, colleges admit students based on their apparent academic merit \( \hat{S}_i \) rather than their actual \( S_i \).

\[
\hat{S}_i = \gamma_x \hat{X}_i + \gamma_y \hat{Y}_i + \gamma_z \hat{Z}_i
\]

The college admits the \( M \) students with the highest \( \hat{S}_i \) values. The college admission problem can be re-expressed as the college admission board selecting the \( \gamma = \{\gamma_x, \gamma_y, \gamma_z\} \) that “best” admits applicants with the highest (actual) academic merit \( S_i \). \(^8\)

Here, we define “best” using the mean-squared error (MSE) loss function: \( \gamma \) is chosen to minimize the expected squared difference between apparent and actual academic merit

\[
\gamma = \arg\min E[ (\hat{S}_i - S_i)^2 ] = \arg\min E_\Omega E_{\hat{\Omega}|\Omega}[ (\hat{S}_i - S_i)^2 ] \quad \text{by Law of Iterated Expectations}
\]

The MSE loss function weighs errors symmetrically and increasingly penalizes larger errors. MSE seems to be a natural loss function for our context as what matters to the college is the relative ranking of applicants. Positive and negative errors, generally, impact relative rankings similarly.

---

\(^3\)These are the qualities most cited by selective colleges. Standardization indicates mean zero and unit variance

\(^8\)This is admittedly a gross simplification of the admission process. For example, the model abstracts away admission criteria focused on recruiting a diverse student body. But this model allows us to illustrate, simply, our key insights, which we believe are robust to the adoption of a more “realistic” admission model
and large errors are much more likely to distort the relative ranking than small ones.\(^9\)

Without loss of generality, we can decompose the applicant signals into information \(\Omega_i\) and noise \(\epsilon_i \equiv \{\epsilon_{xi}, \epsilon_{yi}, \epsilon_{zi}\}\).

\[
\hat{\Omega} = \Omega_i + \epsilon_i = \left\{ \begin{array}{c} X_i + \epsilon_{xi} \\ Y_i + \epsilon_{yi} \\ Z_i + \epsilon_{zi} \end{array} \right\}
\]

(4)

The prediction error associated with a given admission criterion \(\gamma\) is

\[
\begin{align*}
\text{Prediction Error} &= (\hat{S}_i - S_i) \\
&= (\gamma_x - \omega_x) X_i + (\gamma_y - \omega_y) Y_i + (\gamma_z - \omega_z) Z_i + \epsilon_{\text{si}}
\end{align*}
\]

(5)

where \(\epsilon_{\text{si}} \equiv \gamma_x \epsilon_{xi} + \gamma_y \epsilon_{yi} + \gamma_z \epsilon_{zi}\)

(6)

\(B_i\) represents the “bias” introduced by the admission standard \(\gamma\) due to the college weighing the signals differently than the underlying characteristics (\(\gamma \neq \omega\)). \(\epsilon_{\text{si}}\) represents the prediction error due to signal noise. Intuitively, a college would introduce more “bias” (\(\uparrow B_i\)) only if it reduced the prediction error due to signal noise (\(\downarrow \epsilon_{\text{si}}\)). So a college chooses a \(\gamma\) different from \(\omega\) in order to account for differences in the noise across the signals.

The college admission problem can now be expressed as

\[
\min \gamma E[ (\hat{S}_i - S_i)^2 ] = E_{\Omega} \left[ E[ (\hat{S}_i - S_i)^2 | \Omega_i ] \right]
\]

\[
= E[ B_i^2 ] + 2 E_{\Omega} \left[ B_i E[ \epsilon_{\text{si}} | \Omega_i ] \right]
\]

Note

\[
B_i \equiv (\gamma_x - \omega_x) X_i + (\gamma_y - \omega_y) Y_i + (\gamma_z - \omega_z) Z_i
\]

\[
\epsilon_{\text{si}} \equiv \gamma_x \epsilon_{xi} + \gamma_y \epsilon_{yi} + \gamma_z \epsilon_{zi}
\]

(7)

The first two moments of the distributions for \(\Omega_i\), \(\epsilon_i\), and \(\epsilon_i | \Omega_i\) are what matter in the college admission problem. This suggests that a large class of observable systematic biases in the signals do not affect the college admission decision in our model.

For example, suppose that \(\epsilon_{xi}\), the noise surrounding the signal for \(X_i\), is known to differ by gender; more specifically, the mean and variance of the signal noise differs by gender: \(\mu_{xm} \neq \mu_{xf}\) and \(\sigma_{xm}^2 \neq \sigma_{xf}^2\). As long as applicant gender and the values of \((\mu_{xm} - \mu_{xf})\) and \(\frac{\sigma_{xm}}{\sigma_{xf}}\) are known, the distortions created by this systematic bias can be eliminated by normalizing the signal:

\[
\hat{X}_i^* = \left\{ \begin{array}{ll} \hat{X}_i & \text{if } i \text{ is male (m)} \\ (\hat{X}_i + \mu_{xm} - \mu_{xf}) \times \left( \frac{\sigma_{xm}}{\sigma_{xf}} \right) & \text{if } i \text{ is female (f)} \end{array} \right.
\]

By using \(\hat{X}_i^*\) instead of \(\hat{X}_i\), the college can eliminate ex ante distortions in the ranking of the applicant due to gender bias in \(\epsilon_{xi}\). The normalization eliminates differences in the relevant first and second moments due to gender bias, thus preserving the ex ante ranking of applicants.

---

\(^9\)Alternatively, we could define the loss function around \(S_i\) rank order. This introduces sizable analytic difficulty
This result is consistent with anecdotal evidence of college admission boards “adjusting” the SAT scores of disadvantaged applicants: “Applicant A scored a 1900 on the SAT but, given her background, that is equivalent to a 2000 by our more standard applicants.” Or, “SATs are less precise signals (higher variance) for lower income students – so we should value SATs less and other factors more for such students.” But such adjustments are possible only if the systematic bias is observable. Referring to the above example, if applicant gender is unobservable, then the college cannot use $\hat{X}_i^*$ instead of $\hat{X}_i$ as the college does not know which of the applicant signals to adjust.

This is precisely why SAT Prep perplexes college admission boards. If the college knew which applicants were “prepped,” the college could adjust SAT scores accordingly, eliminating any “prepped” distortions. However, this is also the reason why applicants prefer to keep their participation hidden. In the absence of a credible mechanism by which to signal the lack of SAT Prep, the college has difficulty separating the high merit “un-prepped” applicants from lower merit “prepped” applicants. In the next section, we explore how this asymmetric information problem, generated by hidden SAT Prep, differs depending on whether score improvements from SAT coaching are uniform or discriminatory and whether participation in SAT coaching is random or endogenous.

3 SAT Coaching and Asymmetric Information

For expositional simplicity, we assume that, absent hidden actions by the applicants, the error surrounding each signal is simply white noise ("measurement error"). Without SAT Coaching,

- Signal noise and applicant characteristics are independent of each other: $\epsilon_i \perp \Omega_i$
- Signals are unbiased: $E(\epsilon_i) = 0$ and $E(\hat{\Omega}_i) = \Omega_i$
- Signal errors are independent of each other: $\epsilon_{xi} \perp \epsilon_{yi}, \epsilon_{zi}$, $\epsilon_{yi} \perp \epsilon_{zi}$

This simplifies the admission objective of the college to selecting $\gamma$ that minimizes:

$$E[ (\hat{S}_i - S_i)^2 ] = E[ B_i^2 ] + E[ \epsilon_{si}^2 ] + 2E[ B_i E[ \epsilon_{si} | \Omega_i ] ] = E[ B_i^2 ] + E[ \epsilon_{si}^2 ] \quad (\star)$$

In the absence of SAT Coaching, $(\star) = 0$ as $E(\epsilon_{si} | \Omega_i) = E(\epsilon_{si}) = 0$ given the above assumptions.\textsuperscript{10} We make one further assumption for expositional purposes:

- Signal noise for $\hat{X}_i$ is lower than for $\hat{Y}_i$ or $\hat{Z}_i$: $\text{Var}(\epsilon_{xi}) < \min\{ \text{Var}(\epsilon_{yi}), \text{Var}(\epsilon_{zi}) \}$

This is motivated by anecdotal evidence of college admission officers viewing SAT scores as “more precise” (but still noisy) signals than high school transcripts (achievement signal) and letters of recommendation (character signal). The assumption implies that any variance trade-off across signals will lead to the SAT signal being emphasized over the other two signals.\textsuperscript{11}

The college admission board selects $\gamma$ to minimize the expected square difference of the prediction error. Given well defined first and second moments (from the college’s belief about

\textsuperscript{10}Later, we discuss how “SAT Coaching” affects the value of $(\star)$.
\textsuperscript{11}The assumption is for expositional simplicity and does not alter the main qualitative results.
the joint distribution of applicant merit characteristics and signal noise), the first order conditions below are both necessary and sufficient as the underlying optimization problem is concave in $\gamma$.

\[
\frac{\partial E[(\hat{S}_i - S_i)^2]}{\partial \gamma_x} = 0 \quad \frac{\partial E[(\hat{S}_i - S_i)^2]}{\partial \gamma_y} = 0 \quad \frac{\partial E[(\hat{S}_i - S_i)^2]}{\partial \gamma_z} = 0
\]  

(9)

The first order conditions can be re-worked to express $\gamma_x$ as a function of $(\gamma_y, \gamma_z)$:

\[
\gamma_x^*(\gamma_y, \gamma_z) = \omega_x \left( \frac{E[\hat{X}_i X_i]}{E[\hat{X}_i^2]} + (\omega_y - \gamma_y) \left( \frac{E[\hat{X}_i Y_i]}{E[\hat{X}_i^2]} \right) + (\omega_z - \gamma_z) \left( \frac{E[\hat{X}_i Z_i]}{E[\hat{X}_i^2]} \right) \right)
\]

\[
- \left( \gamma_y E[\hat{X}_i \epsilon_{yi}] + \gamma_z E[\hat{X}_i \epsilon_{zi}] \right)
\]

(10)

and similarly for $(\gamma_y^*, \gamma_z^*)$. $\gamma_x^*$ can be considered the “optimal” $\gamma_x$ choice for a given $(\gamma_y, \gamma_z)$.

3.1 No “SAT Coaching”

In the absence of “SAT Coaching,” the signal noises are simply measurement error and the first order conditions defining $\gamma^*$ can be simplified further:

\[
\gamma_x^*(\gamma_y, \gamma_z) = \omega_x \left( \frac{E(X_i^2)}{E(X_i^2) + E(\epsilon_{xi}^2)} \right)
\]

\[
+ (\omega_y - \gamma_y) \left( \frac{E(X_i Y_i)}{E(X_i^2) + E(\epsilon_{xi}^2)} \right) + (\omega_z - \gamma_z) \left( \frac{E(X_i Z_i)}{E(X_i^2) + E(\epsilon_{zi}^2)} \right)
\]

(11)

and similarly for $(\gamma_y^*, \gamma_z^*)$.

$\gamma_x^*$ reflects the incentives the college faces to increase/decrease the emphasis it places on $\hat{X}_i$ as an admission criterion, relative to the given weights for the other two signals. Consider the case where we fix $\gamma_y = \omega_y$ and $\gamma_z = \omega_z$, leaving us with $\gamma_x^* = \omega_x \left( \frac{E(X_i^2)}{E(X_i^2) + E(\epsilon_{xi}^2)} \right)$. In this situation, the college chooses a $\gamma_x$ different from its preference $\omega_x$ only because of the variance of the noise signal, $E(\epsilon_{xi}^2)$. As the signal degrades (greater noise variance), the college places less emphasis on $\hat{X}_i$ when evaluating applicants.

More generally, the college has an incentive to pick low values of $\gamma$ as $\gamma$ monotonically increases the signal noise variance: $\frac{\partial E(\epsilon_{xi}^2)}{\partial \gamma} > 0$. The college only chooses to increase $\gamma$ when the resulting increase in signal noise variance is less than the decrease in expected bias squared $E(B_i^2)$. This means that the highest value $\gamma$ would take, when characteristics are uninformative about each other, is $\omega$; a $\gamma > \omega$ leads to both greater expected bias squared and greater signal noise variance.

If the characteristics are sufficiently correlated with each other, then the college may increase the weight even beyond its underlying preference ($\gamma > \omega$). If two characteristics are sufficiently

---

12 Discussing the fully solved optimal $\gamma_x$ is messier and less intuitive

13 This trade-off is also at the root of the familiar error-in-variables “attenuation bias” in linear regression models.
correlated, but one has greater signal noise than the other, the college may reduce overall noise by “over” emphasizing the signal with the lower variance and “under” emphasizing the signal with the higher variance. This shift allows the college to reduce signal noise variance without an equivalent increase in expected bias squared. Consider the extreme case where \( X_i = Y_i \) but \( \text{Var}(\epsilon_{xi}) < \text{Var}(\epsilon_{yi}) \). In this case, the college should completely disregard \( \hat{Y}_i \) and have \( X_i \) reflect both \( X_i \) and \( Y_i \). This reduces signal noise variance without increasing expected bias squared.

This is the trade-off, discussed earlier, between prediction error due to systematic bias (\( B_i \)) and prediction error due to overall signal noise (\( \epsilon_{si} \)). By balancing the over-emphasis of \( X_i \) with the under-emphasis of \( Y_i \) (or \( \hat{Z}_i \)), the college can reduce \( \epsilon_{si} \) without incurring too much additional \( B_i \). A college chooses an acceptance criterion (\( \gamma \)) different from their preference for the underlying characteristics (\( \omega \) only to exploit the bias-variance trade-off. Furthermore, given the assumption of lower signal noise for \( X_i \), this indicates that while \( \gamma_x \) may be greater than \( \omega_x \), \( \{\gamma_y, \gamma_z\} \) are unlikely to be greater than \( \{\omega_y, \omega_z\} \) unless \( Y_i \) and \( Z_i \) are highly correlated with each other (so one proxies for the other at a lower signal noise variance).

### 3.2 “SAT Coaching”

Consider now the case where there is potential “SAT Coaching.” With possible SAT Prep, the distribution for \( \tilde{X}_i = X_i + \epsilon_{xi} \) depends on SAT Prep status

\[
\tilde{X}_i = \begin{cases} 
X_i + \epsilon_{xi0} & \text{if “unprepped”} \\
X_i + \epsilon_{x1i} & \text{if “prepped”}
\end{cases}
\]

where \( \epsilon_{xi0} \) and \( \epsilon_{x1i} \) are drawn from separate distributions \( f(\epsilon_{xi0}) \) and \( f(\epsilon_{x1i}) \), respectively. Thus,

\[
f(\epsilon_{xi}) = f(\epsilon_{x1i}) \times \theta_i + f(\epsilon_{xi0}) \times (1 - \theta_i)
\]

where \( \theta_i \) is the probability that applicant \( i \) is “prepped” given her underlying characteristics \( \Omega_i \). We assume that the only difference in the two distributions is in the mean: \( E(\epsilon_{x1i}) > E(\epsilon_{xi0}) = 0 \). SAT Prep, on average, helps applicants achieve higher SAT scores.\(^{14}\)

The above highlights two important dimensions along which SAT Prep can vary. The impact of SAT Prep on SAT signal noise may differ by applicant merit characteristics, as \( f(\epsilon_{x1i}) \) can be discriminatory, \( f(\epsilon_{xi0} | \Omega_i) \neq f(\epsilon_{x1i}) \), or uniform, \( f(\epsilon_{xi0} | \Omega_i) = f(\epsilon_{x1i}) \). And, participation in SAT Prep may either be random (\( \theta_i = \theta \)) or endogenously selective (\( \theta_i \neq \theta \)).

<table>
<thead>
<tr>
<th>Table 1: General Types of “SAT Coaching”</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAT Coaching Participation</strong></td>
</tr>
<tr>
<td><strong>Random</strong></td>
</tr>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>Uniform</td>
</tr>
<tr>
<td>Pr((\text{Coaching})) \perp \Omega_i ((\theta_i = \theta))</td>
</tr>
<tr>
<td>( f(\epsilon_{x1}</td>
</tr>
<tr>
<td><strong>Endogenous</strong></td>
</tr>
<tr>
<td>Pr((\text{Coaching})) \depends on \Omega_i ((\theta_i \neq \theta))</td>
</tr>
<tr>
<td>( f(\epsilon_{x1}</td>
</tr>
</tbody>
</table>

| **Discrim.**                              |
|  Pr(\(\text{Coaching}\)) \perp \Omega_i (\(\theta_i = \theta\))  |
|  \( f(\epsilon_{x1} | \Omega_i) \neq f(\epsilon_{x1}) \)  |
|  Pr(\(\text{Coaching}\)) \depends on \Omega_i (\(\theta_i \neq \theta\))  |
|  \( f(\epsilon_{x1} | \Omega_i) \neq f(\epsilon_{x1}) \)  |

\(^{14}\)We ignore other possible coaching benefits, such as lower variance (insurance). We also assume that coaching does not alter the underlying merit characteristics (\( \Omega_i \)).
Depending on the perceived type of effect and participation, the college’s response to possible SAT Prep will differ. This is mainly because $E(\epsilon_{si} \mid \Omega_i)$ is, generally, no longer zero and varies with the type of coaching effect and participation.

$$E(\hat{S}_i - S_i)^2 = E[B_i^2] + E[\epsilon_{si}^2] + 2 E[ B_i E[ \epsilon_{si} \mid \Omega_i ]]$$

Additionally, the value of $E[\epsilon_{si}^2]$ may change with SAT Prep. As the college’s admission objective has changed, so has the first order condition defining the college’s selection of $\gamma$:

$$\gamma_x^*(\gamma_y, \gamma_z) = \omega_x \left( \frac{E(X_i^2) + E(\epsilon_{xi}X_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right)$$

$$+ \ (\omega_y - \gamma_y) \left( \frac{E(X_iY_i) + E(\epsilon_{yi}Y_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right)$$

$$+ \ (\omega_z - \gamma_z) \left( \frac{E(X_iZ_i) + E(\epsilon_{zi}Z_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right)$$

$$- \ \left( \frac{\gamma_y E(\epsilon_{xi}\epsilon_{yi}) + \gamma_z E(\epsilon_{xi}\epsilon_{zi})}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right)$$

and similarly for $(\gamma_y^*, \gamma_z^*)$.

For simplicity, we assume that SAT Prep is uncorrelated with other signal noises $(\epsilon_{yi}, \epsilon_{zi})$. $E(\epsilon_{xi}\epsilon_{yi}) = E(\epsilon_{xi}\epsilon_{zi}) = 0$ and the last term above disappears.\(^\text{15}\) Each of the two main incentives discussed earlier – choosing a $\gamma_x$ close to $\omega_x$ to reduce prediction error and exploiting the correlation between $X_i$ and the other two applicant characteristics $\{Y_i, Z_i\}$ to reduce overall signal noise – remain but are distorted by the presence of hidden SAT Prep.

### 3.2.1 Incentive to Reduce Prediction Bias

The incentive to choose $\gamma_x$ close to $\omega_x$ is captured by the first term

$$\omega_x \left( \frac{E(X_i^2) + E(\epsilon_{xi}X_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right)$$

The term differs from its analog under no SAT Prep in two manners. First, the term now includes $E[\epsilon_{xi}X_i]$, which is generally non-zero. Second, the value of $E[\epsilon_{xi}^2]$ is different with SAT Prep.

Consider the first distortion. If the variation in $\hat{X}_i$ is mostly information, i.e. $E(X_i^2) > E(\epsilon_{xi}^2)$, then

$$\omega_x \left( \frac{E(X_i^2) + E(\epsilon_{xi}X_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right) < \omega_x \left( \frac{E(X_i^2)}{E(X_i^2) + E(\epsilon_{xi}^2)} \right)$$

if $E(\epsilon_{xi}X_i) > 0$

$$\omega_x \left( \frac{E(X_i^2) + E(\epsilon_{xi}X_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2 E(\epsilon_{xi}X_i)} \right) > \omega_x \left( \frac{E(X_i^2)}{E(X_i^2) + E(\epsilon_{xi}^2)} \right)$$

if $E(\epsilon_{xi}X_i) < 0$

\(^\text{15}\)This abstracts away coaching aimed at helping the applicant’s entire portfolio, “college prep”
In other words, if the variation in the SAT signal is mostly information \( (E(X_i^2) > E(\epsilon_{xi}^2)) \) and signal noise is not too negatively correlated with “aptitude” \( (E(\epsilon_{xi}X_i) > 0) \), the incentive to weight the SAT signal similar to the underlying preference for “aptitude” \( (\gamma_x \to \omega_x) \) is weaker with hidden SAT Prep. But if the variation in the SAT signal is mostly noise or the signal noise highly negatively correlated with “aptitude,” the incentive is stronger with hidden SAT Prep.\(^{16}\)

This difference between the no SAT Prep and hidden SAT Prep scenarios stems from the fact that hidden SAT prep may either be a separating or pooling instrument. \( E(\epsilon_{xi}X_i) = E_{\Omega}(E(\epsilon_{xi} \mid \Omega_i)X_i) \) and \( E(\epsilon_{xi} \mid \Omega_i) = \theta_iE(\epsilon_{x1i} \mid \Omega_i).\)\(^{17}\) Therefore, \( E(\epsilon_{xi}X_i) = E_{\Omega}(\theta_iE(\epsilon_{x1i} \mid \Omega_i)X_i). \) So \( E(\epsilon_{xi}X_i) \) reflects the correlation between \( \theta_iE(\epsilon_{x1i} \mid \Omega_i) \) and \( X_i. \) \( E(\epsilon_{x1i} \mid \Omega_i) \) is the expected bump in SAT score from hidden SAT Prep for a student with \( \Omega_i \) characteristics. So \( \theta_iE(\epsilon_{x1i} \mid \Omega_i) \) is the \textit{ex ante} expected bump in SAT score, reflecting not only how SAT Coaching might help an \( \Omega_i \) student but also the probability that an \( \Omega_i \) student will participate in SAT Coaching. \( E(\epsilon_{xi}X_i) = E(\theta_iE(\epsilon_{x1i})X_i) \) reflects the correlation between the \textit{ex ante} expected bump in SAT score from SAT Prep and \( X_i. \)

\( E(\epsilon_{xi}X_i) \) is positive when the \textit{ex ante} expected bump is not too negatively correlated with \( X_i. \) For the most part, \( E(\epsilon_{xi}X_i) \) is positive when the correlation between \textit{ex ante} expected bump and \( X_i \) is positive, when high aptitude applicants are more likely to benefit from the availability of SAT Coaching than low aptitude applicants. In this situation, SAT Prep leads to high aptitude applicants further separating themselves from their low aptitude rivals. What used to be, say, a 50 point expected SAT score difference between high and low aptitude applicants gets magnified to, say, 100 points with such positively correlated SAT Prep. The SAT score difference now exaggerates the aptitude difference; the college has less of an incentive to emphasize the SAT signal as big score differences may imply much smaller aptitude differences.

\( E(\epsilon_{xi}X_i) \) is negative when the \textit{ex ante} expected bump is strongly negatively correlated with \( X_i \) – low aptitude applicants are more likely to benefit from the availability of SAT Coaching than high aptitude applicants. In this situation, SAT Prep leads to low aptitude applicants further pooling themselves with high aptitude applicants. What used to be, say, a 50 point expected SAT score difference between high and low aptitude applicants gets reduced to, say, 25 points with such negatively correlated SAT Prep. The SAT score difference now understates the aptitude difference; the college has more of an incentive to emphasize the SAT signal as small score differences may imply much larger aptitude differences.

In short, if variation in the SAT signal is sufficiently informative, the college is more eager to reduce prediction bias and choose \( \gamma_x \) more closely with \( \omega_x \) when SAT Coaching is a pooling instrument but less eager when SAT Coaching is a separating instrument. But if the variation in SAT signal is mostly noise, the relationship is reversed: for noisy SAT signals, a separating SAT Coaching makes the college more eager to choose \( \gamma_x \) closer to \( \omega_x \) as the separation helps cut through the signal noise. The key point is that SAT Coaching can either strengthen or weaken the incentive to choose a \( \gamma_x \) closer to \( \omega_x , \) depending on signal quality of SATs and the correlation between the

\(^{16}\)If the variation is mostly noise and the noise around the signal is highly negatively correlated with “aptitude,” then the incentive is, again, weaker with hidden SAT Prep. SATs are noisy and misleading with SAT Prep

\(^{17}\)\( E(\epsilon_{x1i} \mid \Omega_i) = 0 \) given the earlier distributional assumptions
ex ante benefit from SAT Coaching \((E(\epsilon_{xi} \mid \Omega_i))\) and aptitude \((X_i)\).

The second distortion stems from the effect of SAT Prep on \(E[\epsilon_{xi}^2]\). The variance of the SAT signal noise is the same for the two scenarios but the mean is greater under SAT Prep. Thus, \(E[\epsilon_{xi}^2]\) is greater under SAT Prep. Specifically, \(E[\epsilon_{xi}^2]\) increases by the square of the now positive mean, \((E[\epsilon_{xi}])^2\). This raises the denominator and lowers the incentive for the college to choose a \(\gamma_x\) closer to \(\omega_x\). The intuition here is that as the SAT signal, in expectation, over-states the “aptitude” of the applicant, the college has an incentive to reduce the weight it assigns to the SAT in order to avoid over-valuing “aptitude” in its deliberations. Combining the two distortions:

- If the variation in \(\hat{X}_i\) is mostly information \((E[X_i^2] > E[\epsilon_{xi}^2])\) then incentive to weight \(\gamma_x\) closer to \(\omega_x\) is weaker with SAT Prep for \(E[\epsilon_{xi}X_i] < -\frac{E[X_i^2]}{E[X_i^2] - E[\epsilon_{xi}^2]} \cdot (E[\epsilon_{xi}])^2\)

The result is derived from comparing the two terms, under SAT Prep and no SAT Prep, as before but now accounting for the difference in \(E[\epsilon_{xi}^2]\) value. In summary, assuming that the variation in the SAT signal is mostly information, the presence of hidden SAT Prep reduces the incentive of the college to choose \(\gamma_x\) closer to \(\omega_x\) unless SAT Prep is a sufficiently strong pooling instrument, with sufficiency depending on the expected SAT Coaching effect \((E[\epsilon_{xi}])\).

### 3.2.2 Incentive to Reduce Overall Signal Noise

The incentives to exploit the correlation between \(X_i\) and the other two applicant characteristics \(\{Y_i, Z_i\}\) are captured by

\[
(\omega_y - \gamma_y) \left( \frac{E(X_iY_i) + E(\epsilon_{xi}Y_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2E(\epsilon_{xi}X_i)} \right)
\]

\[
(\omega_z - \gamma_z) \left( \frac{E(X_iZ_i) + E(\epsilon_{xi}Z_i)}{E(X_i^2) + E(\epsilon_{xi}^2) + 2E(\epsilon_{xi}X_i)} \right)
\]

The inclusion of \(\{E(\epsilon_{xi}X_i), E(\epsilon_{xi}Y_i), E(\epsilon_{xi}Z_i)\}\) and the earlier discussed increase in \(E[\epsilon_{xi}^2]\) make the terms above different from their analog under no SAT Prep.\(^{18}\)

Greater correlation between \(E(\epsilon_{xi} \mid \Omega_i)\) and \(X_i\) indicates that, for given values of \(E(\epsilon_{xi})\) and \(E(X_i)\), \(E(\epsilon_{xi}X_i)\) is larger, the denominator of the above terms smaller, and the incentive to exploit the correlation among applicant characteristics less. The greater correlation implies that the SAT signal, \(\hat{X}_i\), likely overstates the value of \(X_i\) and, by extension, its value as a proxy for the other applicant characteristics. The college, therefore, de-emphasizes \(\hat{X}_i\) as an indicator of \(\{Y_i, Z_i\}\). When the correlation is sufficiently negative such that \(E(\epsilon_{xi}X_i) < 0\), \(\hat{X}_i\) likely understates its proxy value, leading to the college increasing the additional weight it assigns to \(\hat{X}_i\) to proxy for \(\{Y_i, Z_i\}\).

Greater correlation between \(E(\epsilon_{xi} \mid \Omega_i)\) and \(Y_i\) indicates that, for given values of \(E(\epsilon_{xi} \mid \Omega_i)\) and \(E(Y_i)\), \(E(\epsilon_{xi}Y_i)\) is larger, the numerator of the above term larger, and the incentive to exploit the correlation between \(\hat{X}_i\) and \(Y_i\) more. When participation in and/or efficacy of SAT Prep is

---

\(^{18}\)Under no “SAT prep,” \(E(\epsilon_{xi}Y_i) = E(\epsilon_{x0}Y_i) = E(\epsilon_{x0})E(Y_i) = 0\) and, similarly, \(E(\epsilon_{xi}Z_i) = 0\)
greater with larger $Y_i$, SAT Prep enhances the value of the SAT as a proxy for achievement. SAT Prep that favors higher achieving students, through greater participation and/or efficacy, actually aids the college by expanding the possible bias-signal noise trade-off. But when the correlation between $E(\epsilon_{xi} \mid \Omega_i)$ and $Y_i$ is sufficiently negative such that $E(\epsilon_{xi}Y_i) < 0$, SAT Prep favors lower achieving students and troubles the college by reducing the possible bias-signal noise trade-off. Similar conclusions hold for the correlation between $E(\epsilon_{xi} \mid \Omega_i)$ and $Z_i$.

### Table 2: SAT Prep Bias and $\gamma_x^*$

<table>
<thead>
<tr>
<th>High “Aptitude” ($X_i$)</th>
<th>Favored</th>
<th>Disfavored</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Non-Aptitude ($Y_i, Z_i$)</td>
<td>$\gamma_x^* \uparrow$ if trade-off matters more than signal distortion</td>
<td>$\gamma_x^* \uparrow$ unambiguously</td>
</tr>
<tr>
<td>Favored</td>
<td>$\gamma_x^* \downarrow$ unambiguously</td>
<td>$\gamma_x^* \downarrow$ if trade-off matters more than signal distortion</td>
</tr>
<tr>
<td>Disfavored</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall, SAT Prep provides a disincentive to emphasizing the SAT when it favors applicants with high $X_i$ and an incentive when it favors applicants with high $Y_i$ and $Z_i$. Emphasizing the SAT when SAT Prep favors high aptitude applicants raises the concern that aptitude is excessively favored over the other merit characteristics. But emphasizing the SAT when SAT Prep favors applicant with high non-aptitude merits reduces such concern, as SAT Prep makes high non-aptitude merit applicants more likely to have high SAT scores as well. When SAT Prep favors (or disfavors) all three merit characteristics, the net consequence is ambiguous, depending on whether the relative magnitude of the competing (dis)incentives.

#### 3.2.3 Correlation between $E(\epsilon_{xi} \mid \Omega_i)$ and $\Omega_i$

The correlation between $E(\epsilon_{xi} \mid \Omega_i)$ and $\Omega_i$ can be discussed further by noting that $E(\epsilon_{xi} \mid \Omega_i) = \theta_i E(\epsilon_{xi1} \mid \Omega_i)$. So $E(\epsilon_{xi} \mid \Omega_i) = \theta_i E(\epsilon_{xi1})$. There is no correlation between $E(\epsilon_{xi} \mid \Omega_i)$ and $\Omega_i$: $E[\epsilon_{xi}X_i] = E[\epsilon_{xi}Y_i] = E[\epsilon_{xi}Z_i] = 0$, as the merits are measured in standardized units. The presence of SAT Prep simply reduces the precision of the SAT signal ($E[\epsilon_{xi}^2] \uparrow$), leading the college to de-emphasize the SAT ($\downarrow \gamma_x^*$).

**Random Uniform Coaching:** when participation is random and the coaching effect uniform, $\theta_i = \theta$ and $E(\epsilon_{xi} \mid \Omega_i) = E(\epsilon_{xi})$. So $E(\epsilon_{xi} \mid \Omega_i) = \theta_i E(\epsilon_{xi1})$. There is no correlation between $E(\epsilon_{xi} \mid \Omega_i)$ and $\Omega_i$: $E[\epsilon_{xi}X_i] = E[\epsilon_{xi}Y_i] = E[\epsilon_{xi}Z_i] = 0$, as the merits are measured in standardized units. The presence of SAT Prep simply reduces the precision of the SAT signal ($E[\epsilon_{xi}^2] \uparrow$), leading the college to de-emphasize the SAT ($\downarrow \gamma_x^*$).

**Endogenous Uniform Coaching:** when participation is endogenous and the coaching effect uniform, $\theta_i = \theta_i$ and $E(\epsilon_{xi} \mid \Omega_i) = E(\epsilon_{xi})$. So $E(\epsilon_{xi} \mid \Omega_i) = \theta_i E(\epsilon_{xi1})$. There is only correlation between participation and applicant characteristics. $E(X_i\epsilon_{xi}) = E(\epsilon_{xi1})E(\theta_iX_i)$, $E(Y_i\epsilon_{xi}) = E(\epsilon_{xi1})E(\theta_iY_i)$, and $E(Z_i\epsilon_{xi}) = E(\epsilon_{xi1})E(\theta_iZ_i)$. Whether the college emphasizes or de-emphasizes the SAT depends on the correlation between participation and applicant characteristics. If the college perceives SAT Prep participation as negatively correlated with aptitude and positive with
achievement and character, SAT Prep raises $\gamma_x^*$. 

**Random Discriminatory Coaching:** when participation is random and the coaching effect discriminatory, $\theta_i = \theta$ and $E(\epsilon_{xi} | \Omega_i) = E(\epsilon_{xi} | \Omega_i)$. So $E(\epsilon_{x1i} | \Omega_i) = \theta E(\epsilon_{x1i} | \Omega_i)$. There is only correlation between coaching effect and applicant characteristics. $E(X_i \epsilon_{xi}) = \theta E(\epsilon_{x1i}X_i)$, $E(Y_i \epsilon_{xi}) = \theta E(\epsilon_{x1i}Y_i)$, and $E(Z_i \epsilon_{xi}) = \theta E(\epsilon_{x1i}Z_i)$. Whether the college emphasizes or de-emphasizes the SAT depends on the correlation between coaching effect and applicant characteristics. If the college perceives SAT Prep efficacy as negatively correlated with aptitude and positively with achievement and character, SAT Prep raises $\gamma_x^*$. 

**Endogenous Discriminatory Coaching:** when participation is endogenous and the coaching effect discriminatory, $\theta_i = \theta_i$ and $E(\epsilon_{xi} | \Omega_i) = E(\epsilon_{xi} | \Omega_i)$. So $E(\epsilon_{x1i} | \Omega_i) = E(\theta_i \epsilon_{x1i} | \Omega_i)$. The correlation between $E(\epsilon_{xi} | \Omega_i)$ and $\Omega_i$ cannot be simplified. $E(X_i \epsilon_{xi}) = E(\theta_i \epsilon_{x1i}X_i)$, $E(Y_i \epsilon_{xi}) = E(\theta_i \epsilon_{x1i}Y_i)$, and $E(Z_i \epsilon_{xi}) = E(\theta_i \epsilon_{x1i}Z_i)$. Whether the college emphasizes or de-emphasizes the SAT depends on the correlation between the ex ante bump and applicant characteristics. The net effect of SAT Prep on $\gamma_x^*$ is unambiguous only when SAT Prep participation and efficacy (dis)favor merit characteristics similarly. 

The above suggests that we may investigate the possible impact of SAT Coaching on the college admission process by examining the correlation between participation and applicant characteristics and between coaching effect and applicant characteristics. In the next section, we explore these correlations for clients of a major SAT Prep franchise operating in Orange County, California. We combine this data with public information on the distribution of SAT scores for public high school students from the same school/region. The combination of the two allows us to analyze the correlation between participation and observed student characteristics and between coaching effect and observed student characteristics. 

### 4 Empirical Analysis

Much of the academic empirical literature on SAT Prep has focused on coaching efficacy. Using survey data from samples of SAT takers with self-reported prep status, the studies have sought to estimate the average treatment effect of coaching on final SAT score.\(^{19}\) SAT Prep participation is studied mainly in the context of possible selection bias confounding the estimation of the average treatment effect. Notable recent studies along this line are Hansen (2004) and Domingue & Briggs (2009), both of which use modern matching techniques to estimate average coaching treatment effects with reduced selection concerns. 

The main source of data for these studies are a College Board sponsored national survey of SAT takers in 1995-1996 and recent waves of the National Education Longitudinal Study (NELS). The former was introduced by Powers & Rock (1999) and the latter are widely used longitudinal surveys maintained by the U.S. Department of Education. Each has the advantage of being a nationally representative survey of SAT takers with identified prep status. This contrasts with the data we introduce in this paper. Our primary data comes from the Orange County (CA) operation.

\(^{19}\)See Briggs (2002) for an overview of the early empirical studies on SAT Prep
of a major SAT Prep firm. The data is neither nationally representative nor inclusive of test takers who took no or alternative/competing SAT Prep course.

Our SAT Prep data includes all clients of the studied SAT Prep franchise from May 2006 to April 2007. 1909 clients – juniors and seniors of the 2006-07 academic year – participated in one of 152 offered sessions during the study period. 136 clients were eliminated from our sample due to missing data. Table 2 provides a break down of these clients by high school type and session time:

<table>
<thead>
<tr>
<th>School Type</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA Public</td>
<td>CA Private</td>
</tr>
<tr>
<td>1379</td>
<td>327</td>
</tr>
</tbody>
</table>

We have additional information on 674 of the 1773 clients from a survey we conducted in 116 of the offered sessions. The survey contains self-provided information on the current high school G.P.A. and weekly study habit of the clients. We also asked the degree to which parental influence was a factor in their decision to take the SAT Prep course. For clients who were students of a California public high school, additional data is available from the California Department of Education (CDE). Specifically, the CDE provides the following data for each California public high school: student demographics, state accountability test (API) scores, average SAT and AP scores. The CDE data provides further information about the studied SAT Prep clients and some insight on the SAT performance of the relevant population from which our SAT Prep clients are drawn.

Our data, while limited to clients of a specific SAT Prep operation, does offer some advantages over existing data. First, students in our data experience a fairly homogeneous SAT Prep treatment; they all took the same course offered by the same office. In both the College Board and NELS data, SAT Prep refers to a wide array of coaching options. Second, our data provides additional student information, specifically the high school of attendance and residential zip code. Third, our data reflects coaching geared toward the current version of the SAT. College Board implemented major changes in the SAT in 1994 and again in 2005. Both the Powers & Rock (1999) and NELS data analyze the previous (post-1994 but pre-2005) version of the SAT. These advantages address concerns raised in a recent report commissioned by the National Association for College Admission Counseling (NACAC) that critically reviews the current SAT Prep literature.

Given the pros and cons of our data, we combine our original empirical analysis with relevant

---

20 A small fraction of the observations come from Los Angeles. A different office serves much of Los Angeles.
21 Participation rate was 685/1005 ≈ 68%. We lacked test score data on 11 of the survey participants. The median and mode participation rates across courses were 74% and 100% respectively. The missing responses seem mostly due to instructors failing to distribute the survey in their course.
22 AP tests are subject-specific (e.g. biology) exams administered by the College Board.
23 At the most specific, SAT Prep refers to some commercial course not offered by their school. But even this definition allows for substantial treatment heterogeneity – across firms and across offices within such firms.
24 The changes in 2005 were in response to criticism levied by colleges about the informative value of SAT, most prominently by the University of California system. See Geiser & Studley (2001).
25 See Briggs (2009).
results from existing studies to investigate the correlation between SAT Prep efficacy/participation and applicant merit characteristics.

4.1 Coaching Efficacy

Perhaps the most cited results on the efficacy of SAT Prep are those in Powers & Rock (1999). Powers & Rock (1999) find only modest score improvements stemming from SAT Prep and make the important point that the relevant improvement is not the difference in score before and after coaching for the prepped student but rather the counterfactual difference in final score between the same student receiving and not receiving coaching.\(^{26}\) They note that simply re-taking the SAT can lead to score improvements due to greater test familiarity, a point formally revisited in Vigdor & Clotfelter (2003), and that the true coaching effect is net of such improvements. Briggs (2001) and Domingue & Briggs (2009) find similar coaching effects in their analysis of the NELS data, providing some outside validation for the College Board sponsored results.\(^{27}\)

The main focus of these existing studies is the estimation of an average coaching effect. Some of these studies, most notably Hansen (2004), do explore possible heterogeneity in coaching efficacy, but primarily in terms of race, gender, and social economic status (SES). But differences in coaching efficacy by demographic/SES characteristics is less of a problem for selective college admissions, as such characteristics are directly observable. In some sense, greater coaching efficacy for certain demographics/SES is no different than other sources of SAT advantage such groups may have. The college can adjust the SAT scores of applicants in such groups so that their stochastic properties resemble those of others; the college need not adjust the overall SAT admission weight.

But heterogeneity in coaching efficacy by applicant merit characteristics requires changes in the relative weighting of SAT and other applicant signals, as merit characteristics are not directly observable but explicitly valued in college admission. Existing studies provide some insight in the correlation between coaching efficacy and applicant merit characteristics, mostly through estimated differences in coaching effect by initial SAT score and high school achievement. Viewing the initial SAT score as a proxy for “aptitude” and high school course grades for “achievement,” these estimated differences provide some information on \(E[\epsilon_{xi}X_i]\) and \(E[\epsilon_{xi}Y_i]\).

For one of their empirical models (the “Benson” model), Powers & Rock (1999) is able to investigate heterogeneous coaching effect. They find that initially lower scoring students benefit slightly more from Math coaching and that students with good high school grades benefit slightly more from both Math and Verbal coaching. These results would suggest that selective colleges may benefit from raising the admission weight on SAT \((\gamma_x^* \uparrow)\) as SAT Prep is a pooling instrument (with respect to aptitude) that also helps higher achieving applicants. However, the estimated differential effects are small, suggesting any adjustment should also be small.

Domingue & Briggs (2009) also report some suggestive estimates that higher achieving students may benefit more from SAT Prep. They find that students with Advanced Placement (AP) course experience – indicating a more rigorous course of study – had a coaching effect that was

\(^{26}\) They find coaching to have a net effect of +6 to +12 points on Verbal and +13 to +26 points on Math

\(^{27}\) Even modest coaching effects may be of crucial importance to selective college admissions. See Briggs (2009)
12 points higher, roughly doubling their improvement compared to the average. Under the view, shared by many selective colleges, that a more rigorous course of study indicates higher potential achievement, the AP result suggests that SAT Prep ameliorates the concern that valuing high SAT necessarily detracts from other merits.

The existing evidence on average coaching effect is compelling but, unfortunately, that on heterogeneity in efficacy is less so. Part of the problem stems from existing studies pooling across different SAT Prep treatments; Powers & Rock (1999) provides some preliminary estimates that show coaching efficacy varying noticeably across SAT Prep programs. As the existing studies use samples that have few students employing the same SAT Prep firm, even fewer from the same office, it is difficult for such studies to distinguish efficacy differences due to applicant characteristics from those due to treatment characteristics. Consequently, we explore efficacy differences across applicant characteristics using our data from a single SAT Prep treatment. Our data lacks a proper control group, making it unsuited for the estimation of an average treatment effect. But, under the assumption that the heterogeneity in the counterfactual (non-treatment) score improvement is less than that in the factual (treatment) score improvement, the data may still be informative about the relation between SAT Prep efficacy and applicant merit characteristics.  

We observe the practice exam scores for each client from the studied SAT Prep operation for the 2006-07 academic year. These practice exams are the exams the SAT coaches, themselves, use to gauge student progress. The students take four practice exams during the course. The first exam occurs before any instruction and indicates SAT score before treatment. The final exam occurs at the end of regular instruction and indicates SAT score after treatment. We use the score difference between the first and last exam as our measure of (gross) SAT score improvement. Unfortunately, we do not have the actual SAT scores from these students. But these practice exams are built using questions from actual exams and are considered their statistical equivalents.

Below are the average first and last practice exam score and score improvement for clients who attended an Orange County (OC) high school

\[ \text{Note: Cov}(\Delta SAT_{prep} - \Delta SAT_{no \ prep}, \Omega) = \text{Cov}(\Delta SAT_{prep}, \Omega) - \text{Cov}(\Delta SAT_{no \ prep}, \Omega) \]

\[ \text{Some information is provided by College Board, responding to various “truth in testing” legislation, most notably New York’s Education Testing Act of 1979. Some are acquired more clandestinely, involving SAT Prep instructors “taking” the SAT exams} \]

\[ \text{This measure may be slightly biased upward. A practice engaged by some SAT Prep firms is the use of the “hardest” practice exam as the first and the “easiest” as the last. But the difference in difficulty is considered slight; the firms do not intentionally create an “easy” and “difficult” exam.} \]
Several things stand out from the above table. First, there appears to be sizable average gross score improvements, around 85 points combined for the two traditional SAT sections (Verbal and Math) and an additional 95 points for the new (post-2005) Writing section. Second, the average gross score improvements are highest in Writing and lowest in Verbal. Third, the average gross score improvements are similar across high school type, public and private.

As discussed in Powers & Rock (1999) and Vigdor & Clotfelter (2003), the gross score improvements likely exaggerate the true coaching benefit as simply retaking the SAT can lead to higher scores. Powers & Rock (1999) find that uncoached students who took multiple exams saw score improvements of 21 and 22 points, respectively, for the Verbal and Math sections. This suggests that the proper (average) net score improvement for clients in our data is negligible for Verbal and over 40 points for the Math section, results similar to what Powers & Rock (1999) find for one of the major programs in their limited sub-sample analysis. The finding of Writing and Math being more “coachable” is consistent with our discussions with various SAT Prep professionals.

The similarity in average score improvement between public and private high schools is a result germane to our particular empirical investigation. Selective colleges consider the quality of the high school attended by the applicant in their admission deliberation. While the quality of the high school attended does not necessarily indicate applicant achievement, it is informative about the scope of possible achievement and is, thus, likely positively correlated with applicant achievement. We view high school merit characteristics as being another set of proxies for applicant achievement. Grade Point Average (GPA) is the main achievement proxy used in existing SAT Prep studies. But GPA does not control for differences in academic rigor and grading standards across schools.

The similarity in average score improvement between public and private high school suggests that SAT Prep may have a uniform effect with respect to applicant achievement. But public versus private is a coarse comparison of high school quality; some of the very best public high schools in California, with reputations comparable to leading California private schools, reside in Orange County. For public high schools, we have more detailed measures of school characteristics. In the next section, we exploit this additional information on public high schools to investigate further the possible correlation between SAT Prep efficacy and student achievement.

4.1.1 Empirical Distribution of Coaching Effect

We first consider the empirical distribution of (gross) score improvement by observed “aptitude” and “achievement” categories. The client’s score on the first (initial) SAT practice exam is our best,
albeit noisy, measure of aptitude. For achievement, we consider both school quality measures and the self-reported GPAs. Comparing the empirical distribution across categories for these measures provides some coarse insight on how the distribution of (gross) score improvement varies across merit characteristics. Each figure presents the relevant distributions for a single section of the new SAT: Verbal (Reading), Math, and Writing. The distributions were smoothed using a kernel density, Epanechnikov with bandwidth chosen using the Silverman (1986) rule. Summary statistics for the underlying data can be found in the Appendix.

Figures 1-3 illustrate the empirical distribution of score improvement by initial score group: [i] < 500 [ii] 500 − 600 [iii] ≥ 600. Figure 2, depicting Verbal section score improvement, shows little difference in the distribution across groups; the distributions are unimodal around a similar mean with modest variance difference. But Figures for both the Math and Writing sections show some noticeable differences. Of the two, the difference for the Writing section is more pronounced. For math score improvement, the main difference is in the variance, with the distribution for the lower initial score group displaying a larger variance. But for writing score improvement, there is a sizable shift in the distribution, with the lower initial score group having much more of its frequency in the higher score improvements. Across all three sections, students who have high initial scores (≥ 600) experience a modestly lower average improvement, except possibly for writing.

These figures are consistent with our understanding of SAT Coaching. The writing section is considered the “most coachable” and the verbal section the least. This suggests that we should expect to see the largest difference in coaching effect across initial score group for writing and the least in verbal. Moreover, the pronounced differences in the writing section may be driven in part by the timing of our study; the writing section was introduced into the new SAT shortly before our study. So the “familiarity effect” of Powers & Rock (1998) and Vigdor & Clotfelter (2003) may be more poignant for this new writing section.

Figures 4-6 illustrate the empirical distribution of (gross) score improvement by the 2006 API score of the attended high school: [i] < 800 [ii] 800 − 850 [iii] ≥ 850. API is the mandatory accountability exam administered in all public California high schools. The API score follows closely the quality reputation of the school. The group thresholds were based on the state-wide target of 800: schools in [i] are below, [ii] at or just above and [iii] well above the standard. A bit surprisingly, gross score improvements do not vary with the 2006 API score of the school attended by the client. The distributions are largely the same across the the API groups and exam sections. There is some modest difference in the variance – variance is lower for clients attending higher API schools – but hardly any in the mean. The data does not support any noticeable difference in the distribution of score improvement by school quality, as measured by the 2006 API score.

Lastly, Figures 7-9 illustrate the empirical distribution of (gross) score improvement by self-reported grade point average (GPA), standardized to the usual 4 point scale: [i] ≤ 3.25 [ii] < 3.25 and ≤ 3.75 [iii] > 3.75. The total number of observations used to construct these figures is less than those used to construct the earlier figures, as the sample is limited to clients who completed our survey, as opposed to all clients. The figures depict a rightward shift in the distributions.
(from low to high GPA), indicating that a larger share of clients from the higher GPA group are achieving high score improvements. The result is not unexpected as students who excel in the regular classroom are also likely to excel in the SAT Prep classroom. The result also reinforce the idea that SAT Prep improves the SAT scores of their clients without increasing their underlying academic characteristics. If SAT Prep was compensating for weaker schooling, we would expect to see negative correlation between score improvement and API score and between score improvement and GPA; the room for improvement for students from stronger schools and/or students who have academically achieved more should be less.

The empirical distributions suggest two manners by which coaching efficacy may be discriminatory: [1] lower Writing “aptitude” students may benefit more from SAT Coaching [2] students with higher individual achievement, as measured by GPA, may benefit more from SAT Coaching. But these distributions consider the observed merit measures individually. In the next section, we use linear regression analysis to consider the measures jointly.

4.1.2 Coaching Effect Regressions

We regress the following for all sampled clients who attended an Orange County public high school

\[
\text{Final}_i - \text{Initial}_i = \alpha_0 + \beta_1 \text{Initial}_i + \beta_2 (\text{Initial}_i)^2 + \gamma_1 \text{API}_i + \gamma_2 \text{White}_i + \gamma_3 \text{Meals}_i + \gamma_4 \text{Gifted/Talented}_i \\
+ \gamma_5 \text{English Learners}_i + \gamma_6 \text{College Grads}_i + \gamma_7 \text{Full Cred}_i + \epsilon_i
\]  

(15)

By focusing on public school students, we are able to obtain detailed school characteristics from the California Department of Education (CDE). The client’s score improvement is regressed against a quadratic of her initial score and observed characteristics of her attended school: the 2006 API test score of the school, % of students who are Caucasian, % who qualify for the assisted meals program, % who are in the state sponsored Gifted/Talented program, % who are non-native English speakers, % of students whose parents are college graduates, and % of teachers who have full credentials. The % variables range from 0 to 100.

Table 5: Linear Regression on Score Improvement

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (M+V+W)</th>
<th>Math</th>
<th>Verbal</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coef s.e.</td>
<td>coef s.e.</td>
<td>coef s.e.</td>
<td>coef s.e.</td>
</tr>
<tr>
<td>Constant</td>
<td>-263.7888 198.5761</td>
<td>-26.4537 76.4270</td>
<td>77.0298 81.7384</td>
<td>7.4329 90.8684</td>
</tr>
<tr>
<td>Initial</td>
<td>0.6344** 0.1893</td>
<td>0.6539** 0.1580</td>
<td>-0.5485** 0.2227</td>
<td>0.5596** 0.2223</td>
</tr>
<tr>
<td>(Initial)^2</td>
<td>-0.0002** 0.0001</td>
<td>-0.0007** 0.0001</td>
<td>0.0005** 0.0002</td>
<td>-0.0008** 0.0002</td>
</tr>
<tr>
<td>API 06</td>
<td>0.0847 0.1264</td>
<td>0.0110 0.0588</td>
<td>0.0841 0.0550</td>
<td>0.0614 0.0679</td>
</tr>
<tr>
<td>% White</td>
<td>-0.1689 0.3614</td>
<td>-0.1851 0.1688</td>
<td>0.1055 0.1571</td>
<td>-0.0764 0.1936</td>
</tr>
<tr>
<td>% Meals</td>
<td>1.1499 0.7966</td>
<td>0.6359* 0.3709</td>
<td>0.9234** 0.3458</td>
<td>-0.1798 0.4277</td>
</tr>
<tr>
<td>% Gifted/Talented</td>
<td>0.8003 0.7019</td>
<td>0.1231 0.3276</td>
<td>0.4177 0.3044</td>
<td>0.3813 0.3752</td>
</tr>
<tr>
<td>% English Learners</td>
<td>-2.1329 1.5321</td>
<td>-1.4175* 0.7146</td>
<td>-0.9400 0.6654</td>
<td>0.0784 0.8214</td>
</tr>
<tr>
<td>% College Grads</td>
<td>-0.2432* 0.1327</td>
<td>-0.0033 0.0619</td>
<td>-0.0028 0.0576</td>
<td>-0.2222** 0.0710</td>
</tr>
<tr>
<td>% Full Credential</td>
<td>-0.6340 1.1677</td>
<td>-0.4107 0.5442</td>
<td>0.1662 0.5076</td>
<td>-0.3540 0.6262</td>
</tr>
<tr>
<td>n = 1140</td>
<td>R^2 = 0.024</td>
<td>R^2 = 0.080</td>
<td>R^2 = 0.029</td>
<td>R^2 = 0.107</td>
</tr>
</tbody>
</table>

*p-value ≤ 0.10   **p-value ≤ 0.05
Consider first the relationship between initial score and expected score improvement. The initial score is our best, albeit imperfect, measure of the client’s innate “aptitude.” The coefficients before both the linear and quadratic Initial; terms are statistically significant at conventional significance levels, across all four dependent variables. This would suggest that the expected score gain from “SAT Prep” varies with the underlying “aptitude.” But the table below, which illustrates the combined quadratic effect of initial score on score improvement, mitigates this view.

<table>
<thead>
<tr>
<th>Initial (Initial for Total)</th>
<th>$\beta_1 + \beta_2 (\text{Initial})^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>M</td>
</tr>
<tr>
<td>500 (1500)</td>
<td>502</td>
</tr>
<tr>
<td>600 (1800)</td>
<td>494</td>
</tr>
<tr>
<td>700 (2100)</td>
<td>450</td>
</tr>
</tbody>
</table>

Even for 100 (300 for total) point differences in the initial score, the combined effect on score improvement is largely the same, especially for the verbal section. A formal two-sided t-test of the difference in combined effect cannot reject the null hypothesis of no difference at any conventional significance level among {500 (1500), 600 (1800), 700 (2100)} scores. This suggests that average score improvement is largely unaffected by initial score, except possibly at the highest (700+) scores for the Math and Writing sections.

The estimates also suggest that school characteristics, after conditioning on initial score, do not have a large impact on average score improvement. Of the seven included school characteristics, only three – % Meals, % English Learners, % College Graduates – are statistically significant at the 10% or lower level, the latter only for Writing. Perhaps most surprising was the lack of predictive value of the school’s API score. The estimated effect is neither statistically significant nor substantial; including a quadratic term for API Score does not qualitatively alter results. This implies that “SAT Prep” most likely does not compensate for weaker schooling. But the positive, significant impact of % Meals – around a 6-9 point increase in math/verbal average score improvement for a 10% increase in subsidized meals participation – suggests that “SAT Prep” may modestly compensate for weaknesses in other student achievement inputs (e.g. family resources).

<table>
<thead>
<tr>
<th>Status Type</th>
<th>Verbal (V) First</th>
<th>Verbal (V) Last</th>
<th>Verbal (V) Δ</th>
<th>Math (M) First</th>
<th>Math (M) Last</th>
<th>Math (M) Δ</th>
<th>Writing (W) First</th>
<th>Writing (W) Last</th>
<th>Writing (W) Δ</th>
<th>Total (V+M+W) First</th>
<th>Total (V+M+W) Last</th>
<th>Total (V+M+W) Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public OC</td>
<td>519 (71)</td>
<td>541 (83)</td>
<td>22 (50)</td>
<td>526 (95)</td>
<td>591 (98)</td>
<td>65 (55)</td>
<td>505 (82)</td>
<td>603 (89)</td>
<td>99 (65)</td>
<td>1550 (215)</td>
<td>1736 (236)</td>
<td>186 (115)</td>
</tr>
<tr>
<td>Surveyed</td>
<td>514 (69)</td>
<td>537 (77)</td>
<td>23 (46)</td>
<td>510 (86)</td>
<td>577 (90)</td>
<td>68 (54)</td>
<td>501 (81)</td>
<td>594 (82)</td>
<td>94 (63)</td>
<td>1525 (201)</td>
<td>1709 (215)</td>
<td>184 (108)</td>
</tr>
<tr>
<td>Not Surveyed</td>
<td>521 (73)</td>
<td>543 (86)</td>
<td>22 (52)</td>
<td>533 (98)</td>
<td>597 (101)</td>
<td>64 (56)</td>
<td>506 (83)</td>
<td>607 (91)</td>
<td>101 (65)</td>
<td>1561 (220)</td>
<td>1747 (243)</td>
<td>186 (119)</td>
</tr>
</tbody>
</table>

Numbers in parentheses () are the standard deviation

But observed school characteristics reflect the scope of possible achievement. In contrast, GPA reflects realized achievement, caveat differences in grading standards. For a subset of our studied clients, we observe their self-reported GPA (based on the 4 point system) from a survey

---

32Schooling may still have an impact through the initial score
conducted in some of the 2006-07 sessions. Table 7 compares average testing performance for the surveyed and non-surveyed sub-samples. A better proxy for student achievement may be constructed by combining school characteristics with GPA information. We re-run the earlier regression for just the sub-sample of surveyed clients but include interaction terms between GPA and school characteristics as additional regressors.

Table 8: Linear Regression on Score Improvement, with GPA Interactions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (M+V+W)</th>
<th>Math</th>
<th>Verbal</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coef</td>
<td>s.e.</td>
<td>coef</td>
<td>s.e.</td>
</tr>
<tr>
<td>Constant</td>
<td>-440.7638</td>
<td>2547.221</td>
<td>-525.7203</td>
<td>1249.955</td>
</tr>
<tr>
<td>Initial</td>
<td>-0.1037</td>
<td>0.3584</td>
<td>0.1330</td>
<td>0.2979</td>
</tr>
<tr>
<td>(Initial)$^2$</td>
<td>0.0000</td>
<td>0.0001</td>
<td>-0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>API 06</td>
<td>-6.7212**</td>
<td>2.2936</td>
<td>-1.4706</td>
<td>1.1264</td>
</tr>
<tr>
<td>× GPA</td>
<td>1.8706**</td>
<td>0.6408</td>
<td>0.3963</td>
<td>0.3146</td>
</tr>
<tr>
<td>% White</td>
<td>9.0451</td>
<td>6.3019</td>
<td>5.1906*</td>
<td>3.0956</td>
</tr>
<tr>
<td>× GPA</td>
<td>-2.7774</td>
<td>1.7558</td>
<td>-1.5187*</td>
<td>0.8623</td>
</tr>
<tr>
<td>× GPA</td>
<td>6.0745</td>
<td>4.1895</td>
<td>2.0375</td>
<td>2.0560</td>
</tr>
<tr>
<td>% Gifted/Talented</td>
<td>14.6583</td>
<td>12.0523</td>
<td>9.0244</td>
<td>5.9196</td>
</tr>
<tr>
<td>× GPA</td>
<td>-4.4176</td>
<td>3.3815</td>
<td>-2.7048*</td>
<td>1.6615</td>
</tr>
<tr>
<td>× GPA</td>
<td>-16.2443**</td>
<td>8.2364</td>
<td>-6.7938*</td>
<td>4.0454</td>
</tr>
<tr>
<td>% College Grads</td>
<td>3.2116</td>
<td>2.3186</td>
<td>1.9004*</td>
<td>1.3165</td>
</tr>
<tr>
<td>× GPA</td>
<td>-0.8602</td>
<td>0.6462</td>
<td>-0.4945</td>
<td>0.3169</td>
</tr>
<tr>
<td>% Full Credential</td>
<td>51.9017**</td>
<td>23.8997</td>
<td>11.1873</td>
<td>11.7544</td>
</tr>
<tr>
<td>GPA</td>
<td>275.4649</td>
<td>710.3583</td>
<td>202.1500</td>
<td>348.9680</td>
</tr>
</tbody>
</table>

n = 351 R$^2$ = 0.095 R$^2$ = 0.137 R$^2$ = 0.089 R$^2$ = 0.220

These additional regression results suggest that school characteristics may matter in coaching efficacy, but not in isolation of the student’s actual achievement within the school. API, perhaps the best summary of a California public high school’s quality reputation, matters in the (gross) score improvement regressions when interacted with GPA for the Verbal and Writing sections. SAT Prep efficacy appears greater for students who personally achieve more (higher GPA) but especially those who attend a higher quality school.

Altogether, our data analysis suggests that SAT Prep efficacy is largely uniform with respect to aptitude – with possible exception of the Verbal section – but discriminatory in favor of high achieving students, with achievement measured by a combination of GPA and observable high school characteristics. The results, combined with our earlier more theoretical discussion and focusing only on coaching efficacy, suggests that SAT Prep encourages selective colleges to increase the relative admission weight for the SAT signal as it further pools aptitude without detracting from high achievement as much.

$^{33}$We provide a table in the appendix that compares average school characteristics across the two sub-samples

$^{34}$For the linear regression on the surveyed sub-sample that excludes the GPA interactions, the estimated coefficient before API06 is neither substantial nor statistically significant for any section
4.2 Participation

SAT Coaching participation has received less attention than efficacy in the empirical SAT Prep literature. The analyses on participation are largely summary statistics comparisons of the sampled “prepped” and “un-prepped” students or the estimation of a selection equation, as from the first stage of a Heckman selection model. The applicant characteristics considered in these existing studies are primarily demographic and socio-economic. But some existing studies do consider initial exam score and high school GPA.

Powers & Rock (1999) compare their sampled “coached” and “uncoached” students and find that “coached” students had lower average initial exam scores (PSAT scores) but a greater percentage with a GPA between A- and A+.35 But Domingue & Briggs (2009) obtain different results: “coached” students in their NELS 2002 sample had higher average initial exam scores (PSAT scores) and only slightly higher average GPA (3.09 versus 3.05) and the estimated selection equation suggests that neither initial exam score nor GPA is predictive of SAT Prep participation. Briggs (2002) find similar results using the earlier NELS 1988 sample. Existing studies provide a muddled picture on the relationship between SAT Prep participation and applicant merit characteristics.

Our Orange County SAT Prep data cannot, by itself, be used to examine SAT participation as the data does not include information on students who did not attend the studied SAT Prep program. However, we can glean some information about participation by (1) comparing the “market share” of this SAT Prep program across the public Orange County high schools (2) comparing the SAT performance of the observed clients to those, overall, for students at their attended public high school. The former may inform about how participation varies with school characteristics (“achievement”) and the latter about how participation varies with initial SAT score (“aptitude”).

4.2.1 Market Share

Suppose that participation varies systematically with observable school-wide but not individual student characteristics. This suggests the following model of “SAT Prep” participation

\[ N_{if} = \theta_{if} M_i e^{\epsilon_i} \]
\[ \ln(N_{if}) = \ln(\theta_{if}) + \ln(M_i) + \epsilon_i \quad (16) \]

\( N_{if} \) is the number of students in school \( i \) participating in “SAT Prep” courses offered by firm \( f \), \( M_i \) the total number of SAT takers in school \( i \) (the size of market), and \( \epsilon_i \) some idiosyncratic shock. \( \theta_{if} \) is the parameter that defines firm \( f \)’s market share of school \( i \).

We have data on \( M_i \) for all California public high schools and \( N_{if} \) for all public Orange County high schools and one \( f \). By assuming that the relative market share across firms is constant for the studied high schools, we can simplify the model to

\[ \frac{\ln(N_{if}) - \ln(M_i)}{\ln(N_{if}/M_i)} = \ln(\alpha_f) + \ln(\theta_i) + \epsilon_i = \ln(\alpha_f) + X_i \beta + \epsilon_i \quad (17) \]

35The authors estimate a Heckman style selection model but do not report their selection equation estimates
where $\theta_i = \alpha_f \theta_i, \alpha_f$ the relative market share for firm $f$, and $\theta_i$ the combined market share for all “SAT Prep” firms; the share of students taking no “SAT Prep” at school $i$ is $1 - \theta_i$. $\theta_i$ can be further modeled using covariates. OLS can be applied to the above regression equation to obtain estimates of $\beta_i$, the parameters indicating the degree and type of endogenous “SAT Prep” participation. If the non-constant elements of $\beta_i = 0$ then participation is random. As we only observe $N_{if}$ from one firm, $\alpha_f$ is not identified separately from the constant in $X_{if}$.

The natural log of the surveyed firm’s market share in each of the 52 Orange County public high schools in academic year 2006-07 is regressed against the following school characteristics: the 2006 API test score of the school (measure of school quality), the average number of College Board AP exams taken by a junior or senior (measure of college bound students), % of AP exam scores that are 4 or 5 (measure of school quality), % who are Caucasian, % of students who qualify for the assisted meals program (measure of income), % of students who are in the state sponsored Gifted/Talented program (peer quality), % of students who are non-native English speakers (peer quality), % of students whose parents are college graduates (peer quality), and % of teachers who have full credentials (teacher quality). The % variables range from 0 to 100.

<table>
<thead>
<tr>
<th>Variable</th>
<th>coef</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-8.4266</td>
<td>27.2843</td>
</tr>
<tr>
<td># of APs per Jr/Sr</td>
<td>22.7617***</td>
<td>6.8134</td>
</tr>
<tr>
<td>% of AP Scores 4 or 5</td>
<td>0.0471</td>
<td>0.0424</td>
</tr>
<tr>
<td>API 06</td>
<td>-0.0136</td>
<td>0.0248</td>
</tr>
<tr>
<td>% White</td>
<td>0.0169</td>
<td>0.0842</td>
</tr>
<tr>
<td>% Meals</td>
<td>-0.1694</td>
<td>0.1085</td>
</tr>
<tr>
<td>% Gifted/Talented</td>
<td>-0.1134</td>
<td>0.0850</td>
</tr>
<tr>
<td>% English Learners</td>
<td>0.1060</td>
<td>0.2268</td>
</tr>
<tr>
<td>% College Grads</td>
<td>0.0010</td>
<td>0.0485</td>
</tr>
<tr>
<td>% Full Credential</td>
<td>0.1916</td>
<td>0.1808</td>
</tr>
<tr>
<td>n = 51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2 = 0.534$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The participation regression yields results that are consistent with the earlier coaching effect analysis: “SAT Prep” participation does not appear to vary with most observable school characteristics. The only characteristic with which participation rate seems to vary significantly is the average number of AP exams taken by juniors and seniors. This number indicates the extent to which students at the high school are college bound; so schools with more college bound students have, on average, a larger “SAT Prep” participation rate. But this coefficient most likely does not indicate that higher quality schools have more students participating in “SAT Prep” as none of the other school quality measures, most notably the share of AP exam scores that are 4 and 5 (score required for college credit), are predictive of market share.

The regression above suggests that there are no substantial systematic differences in “SAT Prep” participation across public schools in Orange County. There may be systematic variation across public schools in different U.S. counties, but such variation is likely observed and accounted for by the college admission committee, much like the gender example raised in the earlier theoretical
discussion. It is more difficult for the admission committee to distinguish among schools within the same county, and even more difficult to distinguish among students within the same school. The regression above does not preclude endogenous participation with respect to student characteristics; certain types of students within a school may be more likely to be “prepped.”

4.2.2 Comparison to Overall SAT Performance at Attended High School

An analysis of endogenous participation with respect to student characteristics requires us to compare the distribution of SAT scores for the high school population with that for the surveyed sub-population in our SAT Prep data. Such an involved analysis is outside the scope of this study and may be pursued in a separate, later study. Here, we provide a basic version of such analysis for a single Orange County public high school.36

| Table 10: Comparison for a Single Orange County Public High School (2006-07) |
|-----------------|--------|--------|--------|--------|----------|
| Data Source     | Avg V  | Avg M  | Avg W  | Avg Total| % Total ≥ 1500 |
| Actual (Population, N=411) | 602  | 657   | 602   | 1861   | 86.49   |
| First Practice (SAT Prep Data, N=110) | 526  | 573   | 512   | 1612   | 71.82   |
| Last Practice (SAT Prep Data, N=110)  | 551  | 633   | 601   | 1785   | 84.55   |

The above table shows the average verbal, math, and writing scores and the percentage of total scores ≥ 1500 for students at the chosen school for the 2006-07 academic year. The first row corresponds to the actual SAT scores earned by juniors and seniors at the school. The other two rows correspond to the first and last practice SAT scores earned by juniors and seniors from that school in our SAT Prep data. We have more detailed data for the SAT Prep students but not for the entire school population; the California Department of Education only makes the information in the first row available in their annual SAT report.37

The table suggests that participation is not random. The scores for students at the high school who took our studied SAT Prep course averaged both a first and last practice exam score lower than the actual SAT score average for the entire high school. This is consistent with the idea that, at this high school, SAT Prep is taken primarily by students who, otherwise, would be among the lower scoring members – participation is endogenous, with lower “aptitude” students more likely to participate. While SAT Prep participation does not seem to vary by attended school characteristics, it does seem to vary with the tested SAT “aptitude.”

5 Conclusion

Our empirical analysis suggests that coaching efficacy is discriminatory in favor of high achievement and participation endogenous with respect to the tested aptitude, with lower aptitude students more likely to participate. We also find some evidence that efficacy may be discriminatory against the highest aptitude students. The results, as a whole, imply that SAT Prep is a pooling instrument.

36 The chosen high school is a highly rated “suburban” school, with most of the attending students college bound and many aiming for “Ivies.” More students attended this high school than any other in our sample.

37 For details and the data, see http://www.cde.ca.gov/ds/sp/ai/
that also mitigates the amount that a SAT focus detracts from high achievement. Based on our earlier theoretical discussion, this view of SAT Prep indicates that the college faces some incentive to increase the relative admission weight it places on SAT scores ($\gamma_x \uparrow$), in response to SAT Prep.

This perhaps counterintuitive recommendation may be driven in part by our assumption that SAT Prep only affects the mean of the SAT signal distribution. So a pooling SAT Prep simply re-scales the SAT score, with each point of SAT score difference now representing a larger expected difference in the tested aptitude. If SAT prep also affected the variance, the recommendation may be different. For example, if SAT Prep increased not only the mean but variance of the SAT score of otherwise low scoring students, then the recommendation may be to de-value the SAT as an admission criterion; concerns about the shrinking of the expected score difference may be surpassed by concerns about the increase in noise surrounding the SAT signal. But our SAT Prep data, as summarized in Table 4, does not suggest any variance effect. A comparison of the standard deviation for the first and last practice exams does not exhibit substantial differences, in contrast to a comparison of the mean.\footnote{The verbal section may be an exception.}

The data analysis does raise a puzzle: if coaching effect is largely uniform with respect to aptitude, then why are (otherwise) low scoring students more likely to participate than (otherwise) high scoring students? The nominal cost of coaching is the same for both sets of students. The \textit{ex ante} benefit is also the same, if coaching effect is uniform with respect to the tested underlying “aptitude.” College admission is a largely zero sum game, with admission of one student implying one less available admission for other students. This suggests that, unless the underlying “aptitude” is strongly positively correlated with socio-economic status (economic/opportunity cost of coaching), we would expect participation to seem exogenous with respect to aptitude.

One possible answer is that SAT Prep may significantly affect the value of perceived academic merit, $\hat{S}_i$, but not the rank ordering of such perceived merit across the applicant pool. Without a change in the rank ordering, the actual admission outcome is the same. The change in admission probability effected by SAT Prep may be much higher for otherwise low scoring students, who are more likely to be the marginal applicant. So the real returns to SAT Prep may be higher for otherwise low scoring students. It may be possible to explore this issue \textit{via} Monte Carlo simulations, in lieu of complicated calculations involving order statistics. Monte Carlo simulations can also be used to investigate how differences between the actual and expected (by college) distribution of the SAT signal may distort admission outcome.

Our study takes the distribution of applicant signals as exogenously given. But we would expect such distribution to respond to changes in the college admission criteria. For example, if the college decides to de-emphasize the SAT exam ($\gamma_x \downarrow$), we would expect fewer students to participate in SAT Prep and for the SAT signal distribution to change accordingly. Thus, our data analysis can be best thought as reflecting the equilibrium distribution that arises from the strategic interaction of applicants and college. A better understanding of this equilibrium distribution requires a more strategic model, perhaps one based on the Bayesian Perfect equilibrium concept. We hope to explore such a model in a future paper.
Lastly, we note that SAT Prep is but one possible strategic response, by students, to growing admission selectivity. Moreover, students may not be the only stakeholders responding strategically.\(^{39}\) We believe more research in this general area of strategic response to growing admission selectivity is merited, given the importance of selective colleges as an instrument of socioeconomic mobility and the potential amount of deadweight loss stemming from stakeholders seeking to expand their share of the scarce selective college resource.

References


Domingue, B. and D. Briggs (2009), “Using Linear Regression and Propensity Score Matching to Estimate the Effect of Coaching on the SAT,” mimeo, University of Colorado


\(^{39}\) e.g. Ishii (2010) examines grade inflation as a strategic response by some high schools to admission selectivity


Appendix

**Table A1: Average School Characteristics by Survey Status (Public, Orange County)**

<table>
<thead>
<tr>
<th>Status</th>
<th>API06</th>
<th>% White</th>
<th>% Meals</th>
<th>% Gifted</th>
<th>% Eng Learn</th>
<th>% Col Grads</th>
<th>% Full Cred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public OC</td>
<td>822.45</td>
<td>56.07</td>
<td>10.46</td>
<td>19.62</td>
<td>7.49</td>
<td>36.19</td>
<td>96.87</td>
</tr>
<tr>
<td>n = 1140</td>
<td>(42.29)</td>
<td>(17.36)</td>
<td>(11.82)</td>
<td>(8.47)</td>
<td>(5.87)</td>
<td>(27.44)</td>
<td>(3.44)</td>
</tr>
<tr>
<td>Surveyed</td>
<td>818.83</td>
<td>56.32</td>
<td>11.91</td>
<td>18.66</td>
<td>7.90</td>
<td>33.88</td>
<td>96.91</td>
</tr>
<tr>
<td>n = 351</td>
<td>(41.40)</td>
<td>(17.72)</td>
<td>(13.67)</td>
<td>(8.57)</td>
<td>(6.67)</td>
<td>(26.18)</td>
<td>(3.18)</td>
</tr>
<tr>
<td>Not Surveyed</td>
<td>824.06</td>
<td>55.97</td>
<td>9.82</td>
<td>20.05</td>
<td>7.30</td>
<td>37.21</td>
<td>96.85</td>
</tr>
<tr>
<td>n = 789</td>
<td>(42.60)</td>
<td>(17.20)</td>
<td>(10.84)</td>
<td>(8.40)</td>
<td>(5.48)</td>
<td>(27.94)</td>
<td>(3.55)</td>
</tr>
</tbody>
</table>

Numbers in parentheses () are the standard deviation

**Table A2: Score Improvements by Initial Score**

<table>
<thead>
<tr>
<th>Initial</th>
<th>Obs</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 500</td>
<td>470</td>
<td>72</td>
<td>-110</td>
<td>300</td>
</tr>
<tr>
<td>500 – 600</td>
<td>363</td>
<td>72</td>
<td>-180</td>
<td>230</td>
</tr>
<tr>
<td>≥ 600</td>
<td>307</td>
<td>45</td>
<td>-150</td>
<td>180</td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 500</td>
<td>401</td>
<td>27</td>
<td>-170</td>
<td>230</td>
</tr>
<tr>
<td>500 – 600</td>
<td>572</td>
<td>20</td>
<td>-120</td>
<td>150</td>
</tr>
<tr>
<td>≥ 600</td>
<td>167</td>
<td>18</td>
<td>-100</td>
<td>180</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 500</td>
<td>546</td>
<td>113</td>
<td>-100</td>
<td>320</td>
</tr>
<tr>
<td>500 – 600</td>
<td>412</td>
<td>98</td>
<td>-220</td>
<td>230</td>
</tr>
<tr>
<td>≥ 600</td>
<td>182</td>
<td>58</td>
<td>-140</td>
<td>200</td>
</tr>
</tbody>
</table>

N = 1140 (Orange County, Public High School)

**Table A3: Score Improvements by School API**

<table>
<thead>
<tr>
<th>API 2006</th>
<th>Obs</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>300</td>
<td>68</td>
<td>-100</td>
<td>240</td>
</tr>
<tr>
<td>800 – 850</td>
<td>548</td>
<td>65</td>
<td>-150</td>
<td>300</td>
</tr>
<tr>
<td>≥ 850</td>
<td>292</td>
<td>62</td>
<td>-90</td>
<td>210</td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>300</td>
<td>22</td>
<td>-170</td>
<td>230</td>
</tr>
<tr>
<td>800 – 850</td>
<td>548</td>
<td>20</td>
<td>-170</td>
<td>220</td>
</tr>
<tr>
<td>≥ 850</td>
<td>292</td>
<td>24</td>
<td>-120</td>
<td>200</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>300</td>
<td>95</td>
<td>-70</td>
<td>280</td>
</tr>
<tr>
<td>800 – 850</td>
<td>548</td>
<td>102</td>
<td>-220</td>
<td>320</td>
</tr>
<tr>
<td>≥ 850</td>
<td>292</td>
<td>97</td>
<td>-140</td>
<td>250</td>
</tr>
</tbody>
</table>

N = 1140 (Orange County, Public High School)
Table A4: Score Improvements by Reported GPA

<table>
<thead>
<tr>
<th>GPA</th>
<th>Obs</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math ≤ 3.25</td>
<td>70</td>
<td>60</td>
<td>-70</td>
<td>300</td>
</tr>
<tr>
<td>Math &gt; 3.25 and ≤ 3.75</td>
<td>149</td>
<td>69</td>
<td>-150</td>
<td>220</td>
</tr>
<tr>
<td>Math &gt; 3.75</td>
<td>132</td>
<td>70</td>
<td>-60</td>
<td>210</td>
</tr>
<tr>
<td>Verbal ≤ 3.25</td>
<td>70</td>
<td>22</td>
<td>-90</td>
<td>200</td>
</tr>
<tr>
<td>Verbal &gt; 3.25 and ≤ 3.75</td>
<td>149</td>
<td>19</td>
<td>-90</td>
<td>160</td>
</tr>
<tr>
<td>Verbal &gt; 3.75</td>
<td>132</td>
<td>27</td>
<td>-80</td>
<td>130</td>
</tr>
<tr>
<td>Writing ≤ 3.25</td>
<td>70</td>
<td>86</td>
<td>-50</td>
<td>220</td>
</tr>
<tr>
<td>Writing &gt; 3.25 and ≤ 3.75</td>
<td>149</td>
<td>98</td>
<td>-60</td>
<td>280</td>
</tr>
<tr>
<td>Writing &gt; 3.75</td>
<td>132</td>
<td>93</td>
<td>-70</td>
<td>230</td>
</tr>
</tbody>
</table>

N = 351 (Orange County, Public High School, Surveyed)
Figure 1: Math Improvement by Initial Score

Figure 2: Reading (Verbal) Improvement by Initial Score

Figure 3: Writing Improvement by Initial Score

Figure 4: Math Improvement by School 2006 API Score
Figure 9: Writing Improvement by Student GPA