

NBER WORKING PAPER SERIES

SELECTION BIAS IN COLLEGE ADMISSIONS TEST SCORES

Melissa Clark
Jesse Rothstein
Diane Whitmore Schanzenbach

Working Paper 14265
<http://www.nber.org/papers/w14265>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2008

We thank DeForest McDuff for excellent research assistance and Jeff Smith for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Melissa Clark, Jesse Rothstein, and Diane Whitmore Schanzenbach. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Selection Bias in College Admissions Test Scores
Melissa Clark, Jesse Rothstein, and Diane Whitmore Schanzenbach
NBER Working Paper No. 14265
August 2008
JEL No. C24,I2,J24

ABSTRACT

Data from college admissions tests can provide a valuable measure of student achievement, but the non-representativeness of test-takers is an important concern. We examine selectivity bias in both state-level and school-level SAT and ACT averages. The degree of selectivity may differ importantly across and within schools, and across and within states. To identify within-state selectivity, we use a control function approach that conditions on scores from a representative test. Estimates indicate strong selectivity of test-takers in "ACT states," where most college-bound students take the ACT, and much less selectivity in SAT states. To identify within- and between-school selectivity, we take advantage of a policy reform in Illinois that made taking the ACT a graduation requirement. Estimates based on this policy change indicate substantial positive selection into test participation both across and within schools. Despite this, school-level averages of observed scores are extremely highly correlated with average latent scores, as across-school variation in sample selectivity is small relative to the underlying signal. As a result, in most contexts the use of observed school mean test scores in place of latent means understates the degree of between-school variation in achievement but is otherwise unlikely to lead to misleading conclusions.

Melissa Clark
Mathematica Policy Research
PO Box 2393
Princeton, NJ 08543-2393
mclark@mathematica-mpr.com

Diane Whitmore Schanzenbach
Harris School
University of Chicago
1155 E. 60th Street
Chicago, IL 60637
schanzenbach@uchicago.edu

Jesse Rothstein
Industrial Relations Section
Firestone Library
Princeton University
Princeton, NJ 08544
and NBER
jrothst@princeton.edu

Selection Bias in College Admissions Test Scores

I. Introduction

Data from the two leading college admissions tests—the SAT and the ACT—can provide a valuable measure of student achievement for researchers studying the economics of education. Both the SAT and ACT have been administered in approximately the same forms for several decades, to students from nearly every high school. This permits comparisons between and within states, in cross-section and over time. This is an important advantage over alternative tests administered by individual states or school districts, which change frequently and differ substantially across space.

A central concern about the use of ACT and SAT scores for research is the non-representativeness of test-takers. Both exams are taken primarily by college-bound students, who most likely perform better than would their non-college-bound peers. As a result, comparisons of mean SAT or ACT scores between states or between schools may be misleading about mean latent scores – that is, the average that would be observed were scores available for all students. Previous researchers have concluded that state-level mean SAT scores are substantially biased by selection, and have suggested selection-correction procedures for state-level data (Dynarski, 1987; Dynarski and Gleason, 1993). These rely on strong exclusion restrictions, and fail to make an important distinction between states where the SAT is the dominant exam and those where the ACT is dominant and only the strongest students (who hope to attend out-of-state colleges) take the SAT. There is no evidence available about within-school selection, and therefore little guidance for researchers conducting analyses with test score microdata (see, e.g., Krueger and Whitmore, 2001; Card and Payne, 2002; Rothstein, 2006; Abraham and Clark, 2006; and Hanushek and Taylor, 1990).

We provide new evidence on the selectivity of test participation, with an eye toward the impact of selection on measured mean SAT and ACT scores. We distinguish between selection within and between schools; with substantial sorting of students across schools, the degree of selection into test-taking may differ across these two margins.

We begin by examining state-level test score averages, for which the relevant selectivity combines between- and within-school selection in the state. Dynarski (1987) documents a strong negative across-state correlation between SAT participation rates and mean scores. We show that this largely reflects the contrast between what we call “SAT states,” where most college-bound seniors take the SAT rather than the ACT, and “ACT states,” where traditionally only students hoping to attend elite, private, out-of-state colleges took the SAT. Among SAT states, there is relatively little evidence of selectivity of test participation. However, in ACT states both ACT- and (especially) SAT-takers are quite highly selected. To evaluate the exclusion restriction needed to identify selectivity from cross-sectional data, we also examine a “control function” estimator that uses state-level scores on the NAEP exam, administered to a random sample of students in each state, to absorb potential correlations between test participation rates and average latent achievement. This yields very similar estimates, suggesting that the pure cross-sectional estimator is sufficient to identify the selectivity of test participation.

We then turn to an analysis of selectivity within and between schools, within states. An important concern is that school-level test participation rates may be correlated with the unobserved quality of the school or of its students. Indeed, estimates based on observational variation in test participation yield the implausible result that test-takers are negatively selected within schools, while those based on changes in participation rates over time indicate little selectivity.

These results suggest that an exogenous source of variation in test participation is needed to identify the selectivity parameter of interest. We take advantage of a policy reform in Illinois that substantially increased ACT participation rates but, plausibly, did not have important effects on underlying student achievement. Beginning with the high school class of 2002, the ACT exam was required for graduation; the participation rate in 2002 was 99%, up from 71% in 2001. The impact of the new requirement varied substantially across schools, with larger increases in participation rates in schools that initially had low participation—so were far from compliance—and smaller increases in schools that already participated at high rates. Initial participation rates covaried strongly with the school racial composition, with a strong positive association between the school white share and the participation rate in 2000. This association was substantially attenuated by 2004. Racial composition is thus a strong predictor of the change in participation rates between 2000 and 2004.¹ Under the plausible assumption that the relationship between a school's racial composition and the latent achievement of its students did not change over this four-year period, the former is a valid instrument for the change in participation rates. Results from this analysis indicate strong positive selection into test participation, both within and across schools: Schools with the highest mean latent achievement have the highest participation rates, and within schools those students who would score highest are most likely to write the exam.

In our concluding section, we consider the implications of our estimates for analyses that do not have access to natural experimental variation in test participation. Across-school variation in latent test scores is large relative both to within-school variation and to variation in test participation rates, and is highly correlated with the latter. As a consequence, observed mean scores are nearly perfectly correlated with latent means, though less variable. As a result, school-level analyses that fail to account for selectivity are likely to yield attenuated

¹ It is not, however, a strong predictor in states that did not undergo a policy change.

relationships between test scores and other variables, but if a straightforward correction for this attenuation is made there are unlikely to be large additional biases.

II. Data

Our primary data come from microdata samples of SAT and ACT-takers, matched to the high schools that they attended. Our SAT data set includes 100 percent of test-takers in California and Texas, 100 percent of black and Hispanic test-takers, and a 25 percent random sample of test-takers of other races. The ACT data set includes 50 percent of non-white students and a 25 percent of white students who took the ACT exam.² We focus on school- and state-level mean scores on each exam. Our school-level analyses use data on students who graduated from high school in 1996, 2000, and 2004; we focus on 1996 data for our state-level analysis.

The SAT reports two subscores for each student, math and verbal, and a composite score. Each of the subscores is reported on a scale with mean 500 and standard deviation around 100 among tested students (with a maximum score of 800 and a minimum of 200). We focus on the sum of math and verbal scores, which has mean 1000 and standard deviation around 200. The ACT reports several subscores, including one for a natural science subtest that has no SAT analogue. We focus on the ACT “composite” score that averages over all subjects. We convert this from its native scale (ranging from 1 to 36) to the more familiar SAT scale.³

The SAT and ACT are competing exams, and market shares vary substantially across states. In some states—the Northeast, some of the South, and the Pacific coast—most college-

² In both data sets, minority students were over-sampled to permit more accurate estimation of race-specific models. Due to errors in the processing of the data files, students who failed to report a race are missing from the SAT data (except in California and Texas, where they are included) but are sampled at a 100 percent rate in the ACT data.

³ The concordance comes from Dorans (1999). The correlation between the state-level ACT averages constructed from the student-level data and averages reported by the ACT Corporation is 0.98.

bound students take the SAT, while in others—the Midwest and Plains, and portions of the South—the ACT is more common. Colleges that enroll primarily nearby students (public universities, for example) typically prefer the dominant test in their state, while national, elite universities tend to prefer the SAT but often accept either. When a college will accept either test score, students can sometimes arbitrage between the two, taking both tests and reporting only the higher score.⁴ As a result, students who take the less-popular test in their state—particularly SAT-takers in ACT states—tend to be those who hope to attend the most selective schools.

State-level test participation rates come from *The Digest of Education Statistics* (National Center for Education Statistics, various years) for the SAT and from the ACT Corporation (2003) for the ACT. We define a state as an “ACT state” if its ACT participation rate exceeds 40% and as an “SAT state” if the SAT participation rate exceeds this threshold.⁵ We also use state-level mean scores from the National Assessment of Educational Progress (NAEP; National Center for Education Statistics, 2006) mathematics exam administered to 8th graders in 1992, merged to SAT and ACT data from 1996—the same cohort of students four years later.

Both the SAT and ACT micro data sets contain high school identifiers that can be linked to records from the Common Core of Data (CCD), an annual census of public schools.⁶ From the CCD, we extract the number of 12th graders at the school, the racial composition, and the fraction of students receiving free school lunches. Because sampling rates are so high in both the SAT and ACT data, the simple ratio of the number of (weighted) observations at the school in

⁴ The tests’ content differs somewhat, with more of a focus on knowledge rather than aptitude on the ACT, so a student may select into the test on which he has a comparative advantage. Scores can also be manipulated by re-taking the same exam, as most colleges consider only the highest score (Vigdor and Clotfelter, 2003). Our data report the last recorded score on each exam. We cannot link individual observations between the SAT and ACT data, so are unable to measure how much overlap there is between participation on the two exams.

⁵ No state exceeds both thresholds, while two—Arizona and Nevada—exceed neither and are counted toward neither category. See the online Appendix.

⁶ Our school-level analysis excludes private schools.

our test-taker data sets to the number of 12th graders provides an accurate estimate of the participation rate at all but the smallest schools.⁷ All of our school-level analyses exclude schools with participation rates below 0.02.

Table 1 presents descriptive statistics for the school-level data sets. Average SAT scores are much higher in ACT states than in SAT states. The same is true, to a much lesser degree, for ACT scores in SAT states. Within SAT states, however, schools with high participation rates tend to have higher scores than those with low participation rates, on both the SAT and the ACT. Similarly, within ACT states, schools with high participation rates tend to have higher scores than those with low participation rates on both exams.

III. A Model of Self-Selection and Average Test Scores

College entrance exams are not compulsory, and students who take them incur both financial and time costs – perhaps \$45 for test registration, plus several hours on a Saturday morning to take the exam. For students who do not plan to attend college or who plan to apply only to non-selective colleges that do not require entrance exam scores, there are few benefits to offset these costs. By contrast, students who hope to attend a selective college will likely find that the benefits exceed the costs.

Let μ_{ij} be the net benefit of taking the exam for student i from group j (where a group might be a school or a state). Let θ_j be the mean of this in group j , so that $u_{ij} = \mu_{ij} - \theta_j$ has mean

⁷ Students who take exams prior to 12th grade are coded by the year they attend 12th grade – a student who takes the SAT or ACT as an 11th grader in 2003 is counted toward the 2004 participation measure. The ratio of test takers to 12th grade enrollment overstates the cohort participation rate in schools with large dropout rates. To address this, our denominator for the participation rate is the maximum of the number of 12th graders and the average number of students per grade in grades 9 through 12. The latter average excludes grades with very low enrollment, defined as those with enrollment less than 10% of total enrollment in grades 9-12 at the school. In the schools where still we count more test-takers than students, we set the participation rate to 1 if the difference is less than 10% and drop the school otherwise.

zero in each group. θ_j will tend to be higher in schools where more students are college bound; within schools, u_{ij} will be higher among the college bound than among students not planning to attend college. Both θ_j and u_{ij} may vary across tests as well, and θ_j will generally be higher for the test preferred by local public universities. Students for whom the net benefits of taking the exam are positive ($\mu_{ij} > 0$, or $u_{ij} > -\theta_j$) will take the exam.

Let α_j be the average achievement of students in group j , and let $t_{ij}^* = \alpha_j + \varepsilon_{ij}$ (where ε_{ij} has mean zero in each group) be the latent achievement of student i . For many analytical purposes, it would be useful to observe α_j , and this is the parameter of interest for our investigation of selectivity of test participation. Unfortunately, when test participation is non-random, this is not observed. Instead, the average of observed scores is

$$(1) \quad \bar{t}_j = \alpha_j + E[\varepsilon_{ij} \mid \mu_{ij} > 0] = \alpha_j + E[\varepsilon_{ij} \mid u_{ij} > -\theta_j].$$

The final term here is the mean of ε_{ij} among students for whom the net benefit of test participation is positive. Because benefits and latent achievement are unlikely to be independent – students who hope to attend college probably have higher achievement levels than those who do not – this will in general be non-zero. As a consequence, analyses that take \bar{t}_j as a measure of average achievement in group j may be uninformative about the determinants and consequences of true average achievement, α_j .

If ε_{ij} and u_{ij} are assumed to be bivariate normal with correlation ρ within each group j ,

$$(2) \quad \begin{pmatrix} \varepsilon_{ij} \\ u_{ij} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right),$$

then the group-level participation rate, p_j , indicates the mean test taking propensity in the group:

$p_j = \Pr(\mu_{ij} > 0) = \Pr(u_{ij} > -\theta_j) = \Phi(\theta_j)$, so $\theta_j = \Phi^{-1}(p_j)$. This gives (1) a convenient form:

$$(3) \quad \bar{t}_j = \alpha_j + \rho * \sigma * \lambda(p_j),$$

where $\lambda(p) \equiv \varphi(\Phi^{-1}(p)) / p$ is the Inverse Mills Ratio (hereafter, *IMR*) and $\varphi()$ and $\Phi()$ are the standard normal PDF and CDF, respectively.⁸ Thus, if $\beta = \rho\sigma$ were known, the true mean achievement in group j could be recovered: $\alpha_j = \bar{t}_j - \beta * \lambda(p_j)$. The challenge is in obtaining a consistent estimate of β .⁹

IV. Alternative Empirical Strategies for Identifying the Selectivity of Test Participation

We examine several strategies for identifying within-group selectivity. These differ in their imposed assumptions about the between-school relationship between latent performance and test participation.

A. Cross-sectional analysis

We begin with the case where only a single cross section of test scores is observed. The only available strategy for estimating β is via a regression of group average observed scores on the group-level IMR:

$$(4) \quad \bar{t}_j = \alpha + \lambda(p_j) \beta + ((\alpha_j - \alpha) + \bar{e}_j).$$

This yields a consistent estimate of β if and only if α_j is uncorrelated with $\lambda(p_j)$.¹⁰

⁸ This is a well-known framework, introduced by Gronau (1974). See also Heckman (1979). Normality of student achievement within schools is a plausible assumption. Normality of u is not restrictive: If $F(u)$ is the true CDF of net benefits of test-taking within the school, one can always convert u to a normal distribution with $v = \Phi^{-1}(F(u))$. In this framework, ρ is defined as $\text{corr}(\varepsilon, v)$ rather than $\text{corr}(\varepsilon, u)$. The assumption that ε and u are *bivariate* normal is somewhat more restrictive, but also plausible. We present diagnostic analyses in the Appendix which suggest that the distributional assumption is approximately correct.

⁹ Given β , the two components ρ and σ can be readily distinguished from the within-group variance of observed scores, which can be shown to equal $\sigma^2 - \beta^2 \delta(p_j)$ where $\delta(p) = \lambda(p) * [\lambda(p) + \Phi^{-1}(p)]$. Inference on ρ and σ is somewhat complex. We focus on inference for β ; since $\sigma > 0$, $\beta = 0$ if and only if $\rho = 0$.

¹⁰ Sampling error in p_j is negligible in our data, and the estimates of the sampling error of β reported below neglect it. When we take account of it using the approach proposed by Murphy and Topel (1985), the estimated variance of the coefficients of (4) is essentially unchanged.

This strategy assumes away between-school selectivity: While participation rates may vary across groups, there a group's participation rate (or a monotonic function of it) must be uncorrelated with its average latent achievement. This is not particularly plausible. There are likely unobserved factors that vary across groups and influence both the test participation rate and average performance. For example, suppose that groups are schools, and that one school enrolls academically-strong, college-bound students while another enrolls weaker students who are less likely to want to attend college. The former school will have higher α_j and θ_j (so lower $\lambda(p_j)$) than the latter, producing a downward bias in β and leading to an underestimate of the degree of within-school selectivity of test participation.

B. Control function estimator

If a proxy for latent mean performance is available, bias deriving from non-independence of group-level participation rates and latent achievement can be reduced. We can write:

$$(5) \quad \bar{t}_j = \alpha + \lambda(p_j) \beta + X_j \kappa + ((\alpha_j - X_j \kappa) + \bar{e}_j),$$

where $\alpha + X_j \kappa$ is the best linear predictor of α_j given proxy X_j . The bias in a regression of \bar{t}_j on $\lambda(p_j)$ and X_j depends on the variance of $(\alpha_j - X_j \kappa)$ and on its correlation with $\lambda(p_j)$. The better the proxy, the smaller will be the remaining variance in achievement, and therefore the bias.

C. First-differenced estimates

School-level proxies for latent achievement are rarely available, and other methods are necessary. One option is to take advantage of multiple observations on the selection-test score relationship at the same school from different years. The difference between the expressions for school mean scores given by (4) in $t=0$ and $t=1$ is

$$(6) \quad \bar{t}_{j1} - \bar{t}_{j0} = (\lambda(p_{j1}) - \lambda(p_{j0})) \beta + ((\alpha_{j1} - \alpha_{j0}) + (\bar{e}_{j1} - \bar{e}_{j0})).$$

This first-differenced regression estimates β without bias if

$$(7) \quad \text{corr}(\alpha_{j1} - \alpha_{j0}, \lambda(p_{j1}) - \lambda(p_{j0})) = 0.$$

That is, while there may be a cross-sectional correlation between a school's IMR and its latent average score, there can be no correlation between the *change* in a school's IMR and the *change* in its latent mean score.¹¹ This assumption is somewhat more plausible than the corresponding assumption about levels, but absent an exogenous source of variation in IMRs it is still problematic: School quality (and the quality of a school's students) presumably evolves over time, producing shifts in both α_{jt} and $\lambda(p_{jt})$.

D. First-differenced estimates, with instruments

Let Z_j be a variable that predicts the change in IMRs. If Z_j is uncorrelated with the change in latent performance, it can serve as an instrument in (6), allowing for consistent estimation of β . Below, we form such an instrument by taking advantage of a policy change in Illinois that mandated that all high school graduates starting with the class of 2002 take the ACT. Because demographic characteristics of schools were correlated with pre-implementation participation rates, the "bite" of these policies varied with school demographics. In particular, schools with higher minority shares had lower pre-implementation participation rates and larger increases in participation (and larger reductions in IMRs) surrounding implementation of the policy.¹² Of course, minority share is correlated with mean latent test scores, but it is plausible that the base-year minority share is uncorrelated with the *change* in mean latent scores surrounding the implementation of the policies. If so, this base-year minority share forms a valid instrument for the change in IMRs surrounding the policy implementation.

Of course, the interpretation of the IMR coefficient in a model for average observed scores as a measure of selectivity arises from our assumption that latent achievement and the propensity to write the exam have a bivariate normal distribution within schools (or states). If

¹¹ Note that this approach requires that β not vary with t .

¹² The instrument does not predict changes in test taking rates in states that did not have a policy change. See Appendix Table 1.

this assumption is incorrect, even with fully exogenous variation in $\lambda(\theta_j)$ the coefficient that we obtain will not have a useful interpretation and will not permit a consistent estimate of α_j . In the appendix, we present several informal diagnostics meant to identify important violations of normality. We find no indication that the bivariate normality assumption is importantly violated – polynomials in p_j are not significant predictors of \bar{i}_j conditional on $\lambda(\theta_j)$, and the distribution of observed scores is quite similar to that implied by normality of pre-selection scores.

V. State-Level Analysis

We begin by examining the selectivity of test participation within states. Figure 1A displays the scatterplot of state-level SAT participation rates and mean SAT scores for the 1996 cohort, distinguishing between SAT and ACT states. Consistent with previous analyses by Dynarski (1987) and Dynarski and Gleason (1993), there is a strong inverse relationship between average scores and participation rates. This inverse relationship is driven to a surprising degree, however, by the inclusion of ACT states; within the SAT states the participation-score correlation is much smaller. The graph is thus suggestive of strong positive selection into SAT-taking in the ACT states, as higher participation rates are associated with lower scores, and with weaker selection in SAT states. This is consistent with the idea that in ACT states only high-performing students who hope to attend elite out-of-state colleges take the SAT, while in SAT states all or most college-bound students take the exam. Figure 1B displays the same scatterplot for the ACT (concorded to the SAT scale). The overall negative correlation between test taking rates and average scores shown in the SAT data is not as apparent in the ACT data.

Of course, it is possible that latent ability is correlated with test taking rates, which would affect the observed correlations between test-taking rates and observed scores. For instance, if states with better schools manage to persuade more of their students to attend college, the

correlation between participation and observed scores is inflated, potentially masking a negative causal effect of participation on mean scores arising from positive within-state selection.

To examine this, we turn to state-level NAEP scores, which are from tests taken by random samples of students in each state and are therefore not subject to the selection bias that may be reflected in SAT and ACT averages. Table 2 presents regressions relating state-level score averages to inverse Mills ratios computed from state-level participation rates. The top panel shows results for SAT states, with models for SAT scores in columns 1-3 and models for ACT scores in columns 4-6, and the bottom panel shows parallel results for ACT states.

The estimates in columns 1 and 4 are consistent with the results shown in Figures 1A and 1B: Selection into participation is strongly positive for the SAT in ACT states, and nearly as strong (but much less precisely estimated) for ACT scores in ACT states. Selection is weaker in SAT states, for both exams, and indistinguishable from zero for the ACT. Columns 2 and 5 limit the sample to the set of states for which NAEP data on the math performance of random samples of 1992 8th graders – the cohort that took entrance exams in 1996 – are available. This has little effect, save that the insignificant SAT participation coefficient in SAT states changes sign.

Columns 3 and 6 attempt to absorb the potential correlation between state-level participation rates and mean achievement by controlling for the average performance of the cohort in the state on the 8th grade NAEP math exam in 1992. The inclusion of this control does not dramatically change the point estimates (excepting again SAT scores in SAT states, where the IMR coefficient returns to its original value, still insignificant), though it does dramatically increase the regression goodness-of-fit and the added precision makes the IMR coefficient for ACT scores in ACT states significant. There is clear evidence of positive selection into taking both exams in the ACT states, where the propensity to take each exam is correlated around 0.5

with a student's potential score. Any selection is much weaker in SAT states, where the correlation is estimated around 0.1-0.2 but zero is well within the confidence intervals.

VI. Observational Analyses of Within-School Selection

For many purposes, it is useful to distinguish between across-school variation in test participation rates and within-school selection, each of which can produce selectivity bias in state average scores.¹³ In this section, we present observational analyses that distinguish these two sources of selection, first using cross-sectional analyses of the relationship between school-level test taking rates and average scores, then examining changes over time. Each yields implausible estimates of the key within-school selectivity parameter, ρ , suggesting that the identifying restrictions are violated. In the next section, we exploit the Illinois policy reform to identify within-school selectivity under more plausible assumptions.

A. Cross-sectional estimates

Columns 1-4 of Table 3 present cross-sectional estimates of selectivity in 2000, as in equation (4). The top panel shows results for SAT states, with two specifications estimated on SAT scores in columns 1-2 and the same specifications repeated for ACT scores in columns 3-4, while the bottom panel repeats the analysis for ACT states. The first specification in each pair controls only for state fixed effects, while the second adds controls for the school racial composition and for the fraction of students at the school who are eligible for free school lunches.¹⁴ The inverse Mills ratio coefficients are uniformly negative, though all are reduced

¹³ If the mean latent achievement is independent of the participation rate at the state level, the IMR coefficient from a state-level regression estimates $(\sigma_\alpha^2 + \sigma_\varepsilon^2)^{1/2} \text{corr}(\alpha + \varepsilon, \theta + u) = (\sigma_\theta^2 + \sigma_u^2)^{-1/2} (\sigma_\varepsilon \sigma_u \rho + \sigma_\alpha \sigma_\theta \text{corr}(\alpha, \theta))$.

¹⁴ Results from models that include county fixed-effects are quite similar.

notably by the addition of the demographic controls. The IMR coefficient for ACT scores in SAT states is insignificant and near zero in this specification.

If school participation rates are independent of mean latent scores, the negative relationship between IMRs and average scores indicates negative selection into test participation within schools—that is, that the *weakest* students in each school are the most likely to take the college entrance exams. The ρ estimates are strongly negative for the dominant exam in a state and smaller for the secondary exam. Estimates of the selectivity of SAT-writing in SAT states and of ACT-writing in ACT states are nearly identical. As it is unlikely that test writers are actually negatively selected within schools, we conclude that the assumptions of the model are violated by strong across-school correlations between test-taking rates and mean achievement.

B. Time-differenced estimates

One strategy for removing bias from endogeneity of the school-level participation rate is to focus only on changes over time in school-level participation rates and scores. This strategy removes any fixed components of school quality, and identifies the selectivity parameters if changes in test participation are unrelated to changes in a school's latent score.

Columns 5-8 of Table 3 present estimates of the first-differenced regression, (6). The coefficients on the change in inverse Mills ratio (and estimated ρ 's) are nearly all positive, but quite small. This is particularly true in SAT states, where the estimated ρ s are all less than 0.05. β s and ρ s are somewhat larger in ACT states, but still indicate only minimal selectivity.

The first-differenced specifications do not impose assumptions about the cross-sectional correlation between school mean latent scores (α_{jt}) and propensity to write the exam (θ_{jt}), assuming only that the projection of α_{jt} on $\lambda(p_{jt})$ is the same in the two years. It is thus possible to estimate the between-school selectivity of test participation, which we measure as $\text{corr}(\alpha_{jt}, \theta_{jt})$. This is reported in the last row of each panel. As our restriction that β be constant across years

prevents this correlation from varying substantially over time, we report only the average correlation across 1994 and 2000. This is substantial for the dominant test in each group of states – indicating that better schools have more test writers – but small for the secondary test.

These first-differenced results, while more plausible than those obtained from the purely cross-sectional analysis, cannot be presumed to be consistent. There is no particular reason to think that school quality is constant over this period, nor that changes in quality are for some reason unreflected in participation rates. Changes in quality (or in school composition) would most likely produce a positive correlation between the change in θ and the change in α , biasing the selectivity estimates downward.

VII. The Illinois Policy Reform

To isolate participation rate changes that are exogenous to changes in school quality, we take advantage of a policy change in Illinois, an ACT state. Beginning in Spring 2001, all juniors in Illinois public high schools took the “Prairie State Assessment Examination” (PSAE), which included the ACT exam as one component.¹⁵ The ACT was administered under the same conditions as are used for ordinary ACT administrations, and scores derived from the ACT component of the PSAE are counted as valid for most admissions purposes.

The new requirement led to large increases in test participation in Illinois. Changes in average observed scores accompanying this change in participation rates can identify the selectivity of test participation before the policy change. This requires some care, as changes in IMRs cannot be treated as entirely exogenous to changes in school quality. First, initial test-taking rates varied substantially across schools. As the new requirement had more “bite” in

¹⁵ Students are not required to “pass” the PSAE, and the introduction of the test was not related to any other statewide changes in school policies. Colorado introduced a similar program around the same time, but we focus on Illinois because the sample size is so much larger there.

schools with lower initial participation rates, one might expect that schools with lower quality would have tended to see larger participation increases as they complied with the policy. Any mean reversion in school quality would bias a first-differenced analysis that treated the change in IMRs as exogenous. Second, compliance with the policy was incomplete. The average participation rate, as calculated in our data, rose from 57% in 2000 to 89% in 2004.¹⁶ If schools that complied fully differed in their quality from those that did not, this would again be a source of potential endogeneity of the change in test participation rates.

Fortunately for our analysis, there are clear patterns in the impact of the policy that are plausibly unrelated to changes in school quality. Figure 2 shows the scatterplot of participation rates against the school fraction white for Illinois high schools in 2000, before the Prairie State Assessment Exam was introduced, and in 2004. We distinguish between the Chicago Public Schools and those in other Illinois districts, as the two groups differed substantially on both dimensions in 2000.¹⁷ In the left panel, which uses 2000 data, there is a strong relationship between test participation rates and school racial composition, with much higher participation rates in schools with more white students. This pattern is still present, but much weaker, in 2004.¹⁸ Thus, the increase in participation between 2000 and 2004 was most dramatic for the schools with the lowest white shares.

¹⁶ The ACT Corporation reports that 99% of seniors in 2004 took the exam. The difference likely has to do with the denominator. We calculate the denominator from the Common Core of Data as the greater of the number of 12th graders and the average number of students per grade. If dropout rates are high, the average number of students per grade may over-state the number of potential test takers and lower the test taking rate.

¹⁷ Chicago also implemented a variety of reforms in the years studied that might affect scores, test participation, or the distribution of students across schools. See Jacob (2005). All of our analyses of Illinois data allow for unrestricted differences between Chicago schools and the rest of the state in test score and participation changes between 2000 and 2004.

¹⁸ The basic shape of the data is invariant to the use of the school's 2000 racial composition on the horizontal axis in both years.

We exploit this feature of the data to generate variation in the “bite” of the Prairie State requirement that is plausibly exogenous to changes in school’s latent test scores. Specifically, we use the school fraction black and fraction Hispanic in 2000 as instruments for the change in the school-level IMR between 2000 and 2004.¹⁹ These are valid instruments if the relationship between school racial composition and mean latent score did not change over this time period.²⁰

Table 4 presents several analyses of school ACT averages in Illinois. The first column reports a cross-sectional model using 2000 data. This is quite similar to the analogous model in Table 3, and like that indicates strong negative selection into test participation. The second column replicates the first-differenced specification from Table 3 using changes between 1996 and 2000 and just the Illinois data. Like the earlier first-differenced models, this indicates approximately zero within-school selectivity and strong positive between-school selectivity.

The remaining columns of Table 4 report first-differenced estimates from data spanning the Illinois policy change, relating changes in school mean test scores between 2000 and 2004 to changes in inverse Mills ratios. Column 3 presents OLS estimates. These combine the potentially endogenous variation that identifies the estimates in column 2 with exogenous

¹⁹ Appendix Table 1 presents first-stage regressions of the change in school IMRs on the base-year school racial composition. These have the sign implied by Figure 2. The table also presents a first stage equation estimated using all ACT states *except* Illinois and Colorado (which implemented a similar policy change, but where we have been unable to identify a plausibly exogenous source of variation in the policy’s bite). In this sample, the year-2000 racial composition of the school is not predictive of the 2000-2004 change in the school’s IMR.

²⁰ Data from the NAEP indicate that the black-white mathematics test score gap among Illinois 8th graders grew by 2 points (on a scale with standard deviation 36) between 1990 and 2005, while the Hispanic-white gap fell by 8 points over the same period. If this trend holds equally across and within schools and if there are no spillover effects on white students in high-minority-share schools, we will tend to overstate the within-school selection into participation slightly. A back-of-the-envelope calculation indicates that this might account for as much as one fifth of the ρ estimates that we report below. A second threat to the instruments’ validity is that racial composition in 2000 may be correlated with the change in racial composition over the next four years, which could be reflected in (or could itself reflect) changes in a school’s mean latent score. Specifications reported in Appendix Table 1 control directly for the change in school racial composition, with little effect on the selectivity parameter of interest.

variation deriving from the policy change. The estimates differ dramatically from those in Column 2, indicating a within-school correlation between participation and scores of 0.48. Column 4 presents the IV specification. This indicates that falling IMRs (rising participation rates) produce large reductions in observed average test scores, corresponding to strong positive selectivity ($\rho = 0.75$) of students into test-taking within schools.

We allow the across-school correlation between participation and latent scores – the between-school selectivity of test participation – to vary between 2000 and 2004. Estimates of between-school selectivity are shown at the bottom of Table 4. Those that take advantage of the policy change indicate year-2000 correlations that are notably larger than those from the 1996-2000 analysis. Correlations in 2004 are smaller than in 2000, consistent with the idea that the policy served to reduce the “signal” variation in test participation that might correlate with the quality of the school or its students.

VIII. Implications for Analyses of Unadjusted Means

The final component of our analysis is to draw out the implications of our results for analyses that, because they lack a source of exogenous variation in participation like that produced by the Illinois policy change, are unable to correct for selectivity of test participation. We cannot estimate the precise bias that this produces in analyses of observed averages without producing the exact specification that a researcher might attempt to estimate, but we can report some relevant correlations.

The key parameters for school-level analyses are the correlation between the observed mean score at a school and the underlying latent mean, α , and the relative variances of the two measures. To the extent that the correlation between the measures is large, schools with high observed scores will tend also to have high latent scores, and analyses that treat the observed

mean score as a measure of the school's quality will not be unduly affected by selection bias. That is, coefficient signs will likely be correct; if the variances of the two measures differ, the magnitude of estimated effects will be affected, in predictable ways.

To see this, consider a regression that would be correctly specified if the dependent variable was true mean latent scores α_j , but suppose that the observed mean score \bar{t}_j is used instead. Let R be the coefficient from a projection of \bar{t}_j onto α_j . Then if the true model is:

$$(8) \quad \alpha_j = X_j * \gamma + v_j,$$

we can write (neglecting intercepts)

$$(9) \quad \bar{t}_j = \alpha_j R + (\bar{t}_j - \alpha_j R) = X_j * \gamma * R + v_j + (\bar{t}_j - \alpha_j R).$$

There are two biases introduced by using \bar{t}_j in place of α_j as the dependent variable in (9). First, the X coefficient, γ , is biased by a factor R due to the different scales of the two measures. Second, the regression coefficient will be further biased if $(\bar{t}_j - \alpha_j R)$ is correlated with X . This correlation will vary with the choice of variables to include in X , of course. The resulting bias, however, depends on this correlation as well as on the variance of $(\bar{t}_j - \alpha_j R)$; the bias will necessarily be small if $\text{Var}(\bar{t}_j - \alpha_j R) = \text{Var}(\bar{t}_j) * (1 - (\text{corr}(\bar{t}_j, \alpha_j))^2)$ is small.

Table 5 presents various relevant parameters that are implied by the IV estimates in Table 4. The first row repeats the within-school correlation, ρ . The next several rows show estimates of the across-school standard deviations of average observed scores, mean latent scores, the school-level participation parameter θ , and the selection term in models for observed average scores, $\rho * \sigma * \lambda(\rho)$. All are computed from the same Illinois data used in Table 4. In both 2000 and 2004, the across-school standard deviation of mean observed scores is about 112 SAT points. This is notably smaller than the standard deviation of latent scores, which is around 150 in 2000 and 130 in 2004; comparing this to the within-school standard deviation (σ) of around

185 indicates that about 35% of the variance of latent test scores is between schools. By comparison, the standard deviation of θ is around 0.6. As the within-school variation in test participation is normalized to have variance one, this indicates that only about one quarter of the variation in test-taking propensity is between schools.

The next row displays the overall selectivity of test participation, combining the across- and within-school components. The estimate from the pre-reform period is 0.76, suggesting that test takers are strongly positively selected on the whole. This is the relevant parameter for correcting state mean scores, supposing that the relative importance of within- and between-school components of selectivity do not vary across states. It is somewhat larger than the estimate obtained for all ACT states from our control function approach in Table 2.

The next rows of Table 5 show the correlation between observed and latent scores, first unconditionally and then conditional on the school racial composition. Unconditional correlations are 0.97 and 0.98 in the two years. Conditional correlations are somewhat lower, but remain well above 0.9. With such high correlations, $(1 - \text{corr}(\bar{t}_{jt}, \alpha_j))^2$ is quite small, implying that the omitted variable introduced by the use of observed means rather than latent means, $(\bar{t}_{jt} - \alpha_j R)$, has a standard deviation of only about 20 points. Bias from the omission of this variable is likely to be negligible.

The final row of the table shows R , the coefficient of a projection of latent means onto observed means. This is 0.74 in 2000 and 0.86 in 2004. Thus, coefficients in regressions like (9) will be attenuated relative to the true β by 14-26 percent, with the larger number more relevant in situations where test taking is optional.

Of course, all of these estimates derive from models estimated from data on Illinois ACT-takers. They are dependent on the validity of the exclusion restrictions underlying our IV estimates, and even if these are valid, they may not generalize to other states or to SAT scores.

On the other hand, several patterns in the data suggests that the basic conclusion—that selection bias is unlikely to be an important determinant of the qualitative results of analyses that take school-mean test scores as the dependent variable—is likely to be quite robust. Because test-taking rates are less variable than observed scores and because the two are so highly correlated, observed test scores are unlikely to be seriously misleading about rankings of states or schools even if selection into test participation is much more extreme than our estimates indicate. As a final exercise, we computed latent scores under various assumptions about the selection coefficient, ρ . Table 6 reports summaries of the association between these latent scores and observed school means in 2000, limiting the sample to schools with at least 50 students. The first column reports the correlation between α_j and \bar{t}_j . The second column reports the coefficient of a regression of the latter on the former. The third and fourth columns repeat these measures, after first residualizing both α and \bar{t}_j against the school fraction white.

Regardless of the value of ρ assumed, the correlation between observed and latent school means remains quite high, 0.95 when $\rho = 1$ and higher for smaller values of ρ . Even after residualizing scores against the school racial composition – likely an important control variable in many analyses – these correlations are never lower than 0.88. Regression coefficients are slightly more sensitive, reflecting the dependence of the variance of α on the ρ parameter, but even these are quite stable across a wide range of ρ values around the estimated value of 0.75.

We also explore specifications that vary the comparison group. While the first panel of Table 6 uses all Illinois schools, the second panel limits computation of the statistics to the relatively homogenous subset of schools that serve the Chicago suburbs, and the third panel broadens the sample to all schools in ACT states (demeaning α_j and \bar{t}_j separately for each state). While correlations are somewhat reduced in these samples, all remain extremely high: When we use our estimated ρ , they range from 0.88 to 0.95; even when we assume that $\rho = 1$ the range is

0.80 – 0.92. With such high correlations, there is little room for substantive bias even when test scores are not corrected for selection into test participation.

IX. Conclusion

Researchers often use college admissions test scores as measures of student achievement. But since only a select group of students take these tests, analyses that use observed score averages as a proxy for the mean latent test score may be biased. Correcting the selection bias has been difficult: important unobserved variables likely impact both the propensity to take the test and the latent test score, and as a result standard selection corrections perform poorly.

In this paper, we examine selectivity biases in both state-level and school-level SAT and ACT scores. At the state level, selection appears to introduce important biases only in ACT states; in SAT states the correlation between the propensity to take the test and the latent score is small and the resulting selection bias is also small.

Our school-level analysis uses a policy change in Illinois requiring all high school juniors to take the ACT as a source of plausibly exogenous variation in changes in participation rates over time. We find that test takers are strongly positively selected both across and within schools: Participation rates are higher at higher-performing schools, and within schools higher-achieving students are more likely to take the test. The across-school variation in latest test scores is large relative both to within-school variation in latent scores and to across-school variation in participation rates. As a result, observed mean scores are highly correlated with latent means, and selectivity bias serves to attenuate variations in school mean scores but is unlikely to introduce further bias. Thus, despite the strong selectivity of test participation, simple analyses using observed school-level test scores without controls for selectivity are unlikely to produce misleading results.

It remains for further analysis to investigate the validity of our Illinois-based estimates for data from other states. Based on the patterns seen here, our expectation is that the results will generalize reasonably well to other ACT states. If so, researchers would be justified in using unadjusted ACT averages as measures of student achievement, correcting for attenuation but not further adjusting for selection.

References

- Abraham, Katharine G., and Melissa A. Clark. 2006. "Financial Aid and Students' College Decisions: Evidence from the DC Tuition Assistance Grant Program." *Journal of Human Resources* 41(3): 578-610.
- ACT Corporation. 2003. "ACT Average Composite Scores by State." ACT Corporation website: <http://www.act.org/news/data.html>, May 2003.
- Card, David and A. Abigail Payne. 2002. "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores." *Journal of Public Economics* 83(1): 49-82.
- Dorans, Neil J. 1999. "Correspondences Between ACT and SAT I Scores." *College Board Report No. 99-1*.
- Dynarski, Mark. 1987. "The Scholastic Aptitude Test: Participation and Performance." *Economics of Education Review* 6(3): 263-273.
- Dynarski, Mark and Philip Gleason. 1993. "Using Scholastic Aptitude Test Scores as Indicators of State Educational Performance." *Economics of Education Review* 12(3): 203-211.
- Gronau, Reuben. 1974. "Wage Comparisons--a Selectivity Bias." *The Journal of Political Economy* 82(6): 1119-1143.
- Hanushek, Eric A. and Lori L. Taylor. 1990. "Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement." *The Journal of Human Resources* 25(2): 179-201.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153-161.
- Jacob, Brian. 2005. "Accountability, Incentives, and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics* 89(5-6): 761-796.
- Krueger, Alan B. and Diane M. Whitmore. 2001. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *The Economic Journal* 111(January): 1-28.
- Murphy, Kevin M. and Robert H. Topel. 1985. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business and Economic Statistics* 3(4): 370-379.
- National Center for Education Statistics. Various years. *Digest of Education Statistics*. U.S. Department of Education, Office of Educational Research and Improvement: Washington, DC.
- National Center for Education Statistics. 2006. "NAEP Data Explorer." U.S. Department of Education, National Center for Education Statistics, Institute for Education Sciences website: <http://nces.ed.gov/nationsreportcard/nde/>, accessed April 2006.
- Rothstein, Jesse. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review* 96(4): 1333-1350.
- Vigdor, Jacob L. and Charles T. Clotfelter. 2003. "Retaking the SAT." *Journal of Human Resources* 38(1): 1-33.

Figure 1A. Average 1996 SAT Score versus SAT Participation Rate

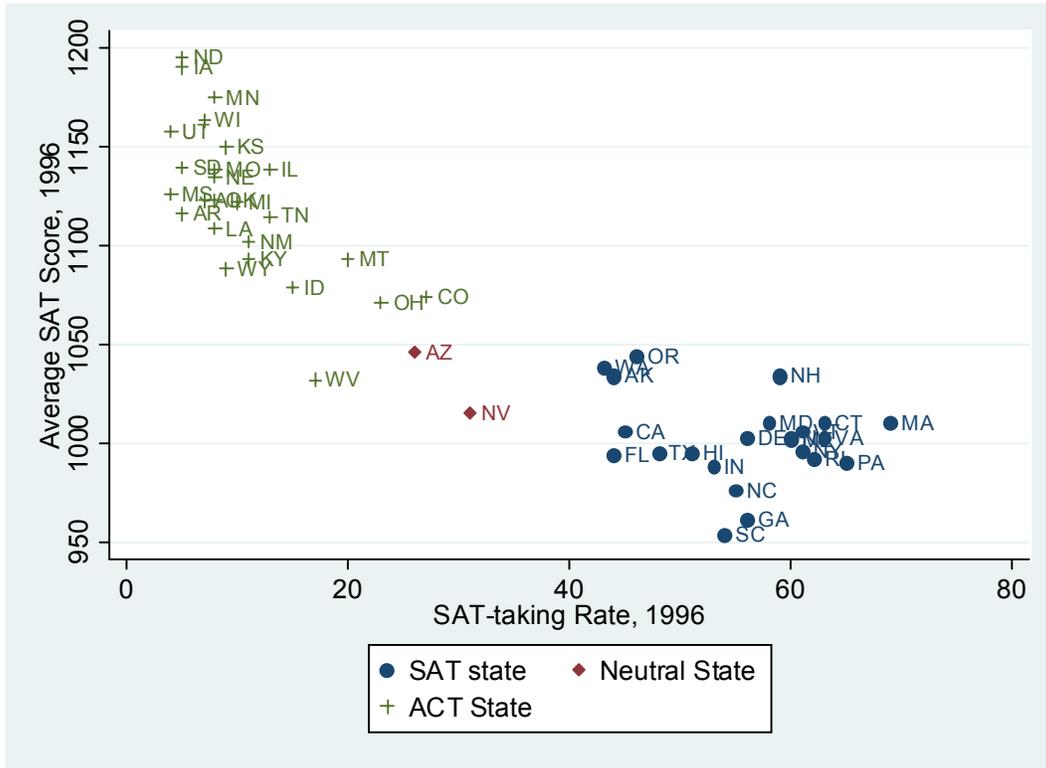


Figure 1B. Average 1996 ACT Score versus ACT Participation Rate

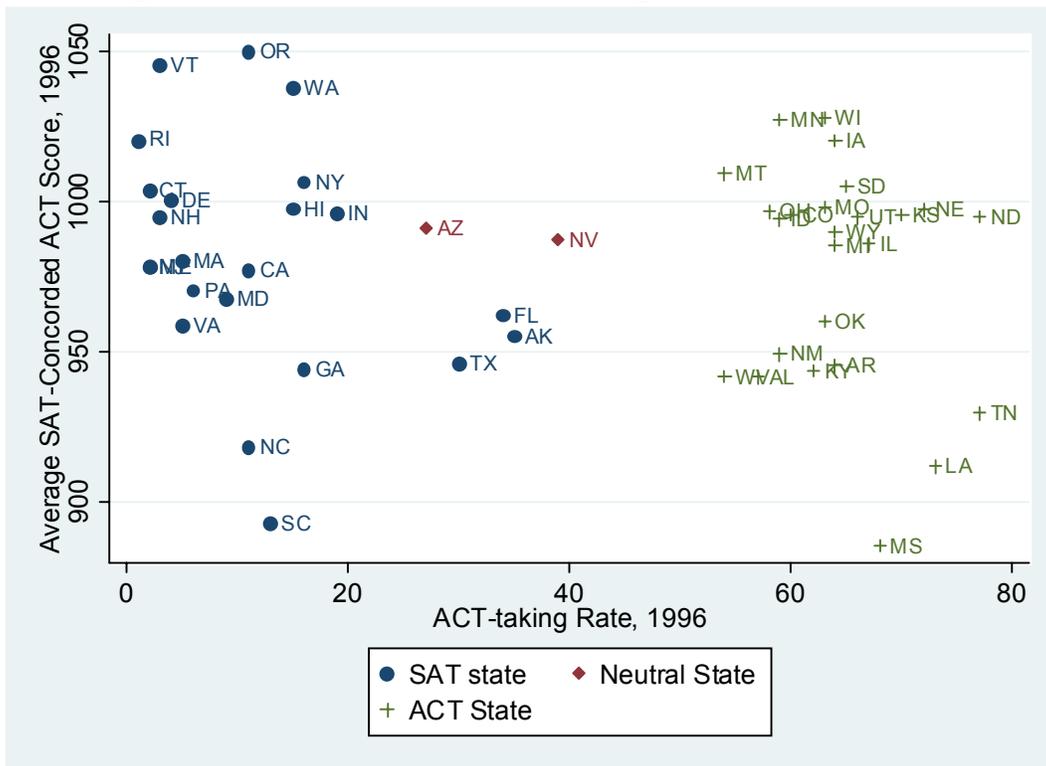


Figure 2. ACT Participation vs. School Racial Composition in Illinois

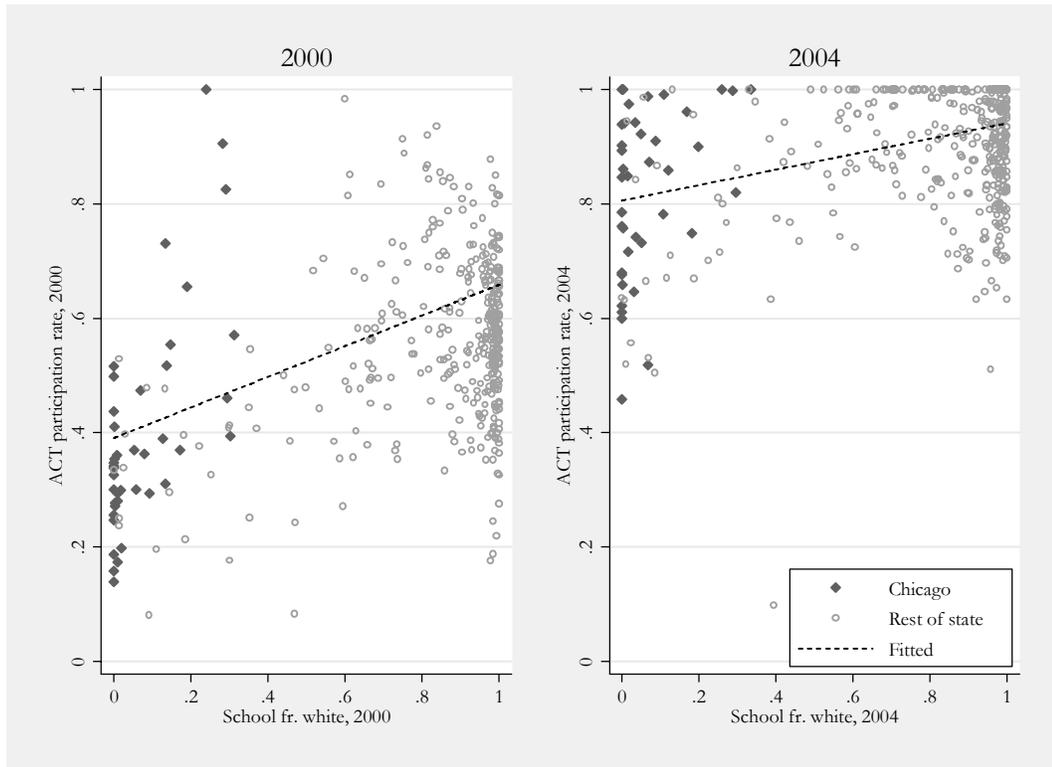


Table 1. Summary statistics

	SAT States				ACT states			
	All schools		Above average participation	Below average participation	All schools		Above average participation	Below average participation
	Mean	S.D.	Mean	Mean	Mean	S.D.	Mean	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SAT								
Mean score	978	100	1037	958	1104	145	1117	1096
School size	285	201	304	279	239	153	270	218
No. of test-takers	130	109	209	104	35	55	65	15
Participation rate	46%	21%	71%	38%	13%	14%	22%	7%
Fr. white	68%	31%	78%	64%	83%	23%	82%	84%
# of schools	5,167		1,291	3,876	1,882		756	1,126
ACT								
Mean score (SAT scale)	970	122	983	962	963	89	999	949
School size	296	202	254	323	174	144	179	173
No. of test-takers	42	49	56	32	99	92	132	86
Participation rate	16%	15%	26%	9%	57%	16%	75%	50%
Fr. white	68%	30%	74%	64%	82%	26%	90%	78%
# of schools	4,143		1,654	2,489	4,050		1,147	2,903

Notes: All data pertain to the 2000 cohort. Sample includes only public schools with participation rates (on the SAT in the first panel or the ACT in the second panel) above 2%. All statistics are computed from unweighted school-level data. Samples in columns 3, 4, 7, and 8 are schools with participation rates on the relevant exam above and below the state-level participation rate. School size is the size of a single grade cohort at the school; see text for details.

Table 2. Cross-sectional estimates of state-level selectivity into test-taking, 1996

	SAT scores vs. SAT participation			ACT scores vs. ACT participation		
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Panel A: SAT States</u>						
Inverse Mills Ratio in test-taking rate in state	45.0 (36.2)	-40.7 (26.7)	38.7 (33.8)	28.4 (10.9)	30.5 (11.9)	21.2 (18.4)
NAEP 8th grade math scores, 1992			2.0 (0.8)			1.0 (1.4)
Exclude states without NAEP data?	n	y	y	n	y	y
N (states)	23	19	19	23	19	19
R2	0.06	0.05	0.39	0.11	0.19	0.21
Underlying parameters						
ρ	0.22	-0.20	0.19	0.15	0.16	0.11
σ	200.4	199.9	199.6	195.0	195.3	194.3
<u>Panel B: ACT States</u>						
Inverse Mills Ratio in test-taking rate in state	110.4 (21.6)	122.0 (26.0)	121.2 (16.4)	77.4 (85.2)	77.5 (96.5)	89.9 (25.7)
NAEP 8th grade math scores, 1992			1.7 (0.4)			3.3 (0.2)
Exclude states without NAEP data?	n	y	y	n	y	y
N (states)	25	21	21	25	21	21
R2	0.52	0.56	0.78	0.04	0.03	0.92
Underlying parameters						
ρ	0.52	0.56	0.55	0.40	0.40	0.46
σ	213.8	219.0	218.6	193.4	193.4	196.3

Note: Average ACT scores are computed from ACT micro-data, concorded to the SAT scale using the concordance developed by Dorans et al. 1999.

Table 3. Cross-sectional and first-differenced estimates of within-school selectivity into test-taking

	Cross-sectional, 2000				First differenced, 1996-2000			
	SAT scores		ACT scores		SAT scores		ACT scores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<u>Panel A: SAT states</u>								
Inverse Mills Ratio in test-taking rate at school, 2000	-170.0 (5.1)	-68.5 (4.7)	-45.0 (5.5)	-2.7 (4.2)				
Change in Inverse Mills Ratio in test-taking rate at school, 1996-2000					4.6 (5.1)	9.3 (4.8)	-0.4 (6.9)	1.2 (6.9)
Controls								
State FEs	y	y	y	y	y	y	y	y
Racial composition, fr. free lunch	n	y	n	y	n	y	n	y
N (schools)	5,167	5,167	4,143	4,143	4,076	4,076	3,122	3,122
R2	0.33	0.65	0.12	0.47	0.01	0.04	0.02	0.03
Underlying parameters								
ρ	-0.75	-0.36	-0.25	-0.02	0.03	0.05	0.00	0.01
σ	226.6	189.3	177.4	172.7	181.8	182.0	174.2	174.2
$\text{corr}(\alpha, \theta)$	0	0	0	0	0.49	0.50	0.08	0.08
<u>Panel B: ACT states</u>								
Inverse Mills Ratio in test-taking rate at school, 2000	-25.0 (6.3)	-19.5 (5.8)	-158.6 (6.4)	-63.8 (5.3)				
Change in Inverse Mills Ratio in test-taking rate at school, 1996-2000					18.2 (14.0)	20.0 (14.1)	19.4 (6.8)	22.3 (6.8)
Controls								
State FEs	y	y	y	y	y	y	y	y
Racial composition, fr. free lunch	n	y	n	y	n	y	n	y
N (schools)	1,882	1,882	4,050	4,050	1,132	1,132	3,286	3,286
R2	0.14	0.34	0.44	0.66	0.01	0.02	0.01	0.03
Underlying parameters								
ρ	-0.15	-0.11	-0.75	-0.36	0.11	0.12	0.11	0.13
σ	171.2	170.6	211.1	179.6	172.0	172.1	173.0	173.2
$\text{corr}(\alpha, \theta)$	0	0	0	0	-0.04	-0.03	0.48	0.49

Table 4. Selection estimates from Illinois data

	Cross-	1st difference,	1st difference, 2000-2004			
	sectional, 2000	1996-2000	OLS	IV	OLS	IV
	OLS	OLS	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Inverse Mills Ratio in test-taking rate at school	-194.6 (16.8)					
Change in Inverse Mills Ratio in test-taking rate at school		8.6 (15.6)	88.0 (12.5)	149.4 (37.5)		
Chicago	-160.1 (14.1)	12.2 (5.3)	57.7 (7.0)	67.8 (11.8)	59.6 (11.3)	75.9 (15.9)
Change in test participation rate at school					-140.0 (22.4)	-302.6 (107.4)
N (schools)	390	390	390	390	390	390
R2	0.73	0.01	0.23	0.16	0.13	0.01
ρ	-0.85	0.05	0.48	0.75		
σ	228.7	174.9	183.4	199.0		
corr(α , θ), 2000	-0.05	0.67	0.75	0.80		
corr(α , θ), 2004			0.48	0.54		

Notes: Sample includes public schools with ACT participation rates above 2% and non-missing demographic information in each of 1996, 2000, and 2004. Instruments in columns 4 and 6 are the fraction black and fraction Hispanic at the school in 2000.

Table 5. Relevant statistics for studies with imperfect selectivity controls implied by the Illinois analysis

	2000	2004
	(1)	(2)
Within-school, individual-level correlation (ρ)	0.75	
Standard deviation across schools		
\bar{t}	112.3	114.7
α	147.8	130.3
θ	0.54	0.60
$\rho^* \sigma^* \lambda(p)$	45.8	28.3
Net correlation (unconditional)		
$\text{corr}(\alpha + \varepsilon, \theta + u)$	0.76	0.68
Across-school correlations		
$\text{corr}(\bar{t}, \alpha)$, unconditional	0.97	0.98
$\text{corr}(\bar{t}, \alpha)$, conditional on racial composition	0.93	0.96
Regression of α on \bar{t}	0.74	0.86

Notes: All entries computed from the model reported in Column 5 of Table 4.

Table 6. Sensitivity of $\bar{\alpha}$ relationship to ρ .

	Scores		Residual scores	
	Correlatio	Reg.	Correlation	Reg. coeff.
	(1)	(2)	(3)	(4)
All schools in Illinois (N=344)				
$\rho = 0$	1	1	1	1
$\rho = 0.25$	1.00	0.91	0.99	0.89
$\rho = 0.5$	0.99	0.81	0.97	0.77
$\rho = 0.6$	0.98	0.77	0.95	0.71
$\rho = 0.75$	0.97	0.71	0.93	0.63
$\rho = 0.9$	0.96	0.64	0.90	0.55
$\rho = 1$	0.95	0.59	0.88	0.49
Schools in chicago suburbs				
$\rho = 0$	1	1	1	1
$\rho = 0.25$	0.99	0.87	0.98	0.87
$\rho = 0.5$	0.98	0.74	0.94	0.71
$\rho = 0.6$	0.97	0.69	0.91	0.65
$\rho = 0.75$	0.95	0.61	0.88	0.55
$\rho = 0.9$	0.93	0.54	0.83	0.46
$\rho = 1$	0.92	0.49	0.80	0.40
All schools in ACT states				
$\rho = 0$	1	1	1	1
$\rho = 0.25$	0.99	0.91	0.99	0.91
$\rho = 0.5$	0.98	0.80	0.95	0.78
$\rho = 0.6$	0.97	0.75	0.93	0.72
$\rho = 0.75$	0.95	0.68	0.90	0.63
$\rho = 0.9$	0.92	0.61	0.86	0.54
$\rho = 1$	0.91	0.55	0.83	0.48

Notes: Schools with fewer than 50 students per cohort are excluded. "Residual scores" are residuals from regressions of mean scores and latent scores on the school fraction white. The sample for these regressions is Illinois schools in the first two panels; all schools in ACT states (with state fixed effects) in the last panel. Correlation and regression coefficients in the final panel are based on variables that have had a state mean removed.