A Test of "Fair Treatment" with Application to University Admissions

Debopam Bhattacharya, Shin Kanaya, Margaret Stevens
University of Oxford

September, 2011.
(Preliminary Draft)

Abstract

Selective colleges are often criticized for using "unfair" admission policies that favour applicants from specific socioeconomic groups. In this paper, we develop a statistical framework for testing whether current admission practices maximize expected academic performance of the entrants, using data on admissions. Such maximization would imply that the expected performance of the marginal admitted candidate –the admission threshold– is equalized across socioeconomic groups. For biased practices, the marginal candidate in the discriminated group will have higher expected outcomes. When we observe for every applicant all of his/her pre-admission characteristics systematically used by the college and post admission performance of admitted candidates, the admission threshold can be identified as long as the density of expected performance is positive around that threshold. This leads to a test of whether the observed socioeconomic disparities among the admitted cohort can be justified as the consequence of admitting candidates with the highest academic potential. If the admission thresholds are found to differ, say, due to affirmative action policies, then we propose a method of calculating the resulting efficiency loss, in terms of academic achievement foregone. Our identification strategy leads to natural estimators and we derive the relevant large sample distribution theory on which to base inference. Applying these methods to two cohorts of UK and EU-based applications to Oxford's flagship PPE degree, we find that male applicants face a higher admission threshold than female applicants and that there is no difference between thresholds faced by state-school and private school pupils.

Keywords: University admissions, discrimination, selection on observables, conditional median restriction.

1 Introduction

Every year, thousands of students around the world apply for admission to colleges and universities. The colleges' decisions on whom to admit are typically based on high-school academic performance as well as scores on standardized aptitude tests, such as the ACT and SAT in the US and the Oxford and Cambridge TSA in the UK. In addition, some colleges take into account the applicants' demographic backgrounds and some conduct expensive admission interviews. In this context, it becomes important from a policy perspective to explore two related questions regarding the admissions process – (a) how the observed applicant characteristics and test-scores should be weighed in the admission process and (b) how to test whether prevalent admission procedures are "biased" against applicants from disadvantaged backgrounds.

There is a substantial literature on the first question in educational statistics, using "validity" studies, which analyze the relative importance of aptitude test scores vis-a-vis measures of high school academic performance and/or socioeconomic status (SES, henceforth) in predicting post-enrolment academic outcomes. See, for instance, Kobrin et al (2004), Noble (2003), Parks et al (2011), Sawyer (1996), Sackett et al (2009) and the references therein as well as a large number of staff research papers made available online by the College Board and the act.org in the United States and the Institute of Education in the UK. The generic approach in this literature is to focus on a specific post-enrolment outcome, such as freshman year or final year GPA (or a nondecreasing function of it) after an applicant enters college. Then one runs a regression – typically linear or logistic – of this outcome on one or more pre-enrollment achievement measures, such as high-school GPA and aptitude test scores. However, in most cases, outcomes are observed only for those granted admission and they are obviously not a random sample from the population of all applicants. In such cases, the regression coefficients obtained for the entrants are assumed to be identical for the general population of applicants. One then compares the proportion of outcome variation explained by the aptitude test scores vis-a-vis the part explained by high-school GPA and/or SES to judge the incremental contribution of the aptitude test. In some studies, an "optimal" cut-off rule is derived for the entire applicant population such that if an applicant's regression coefficient weighted sum of test-scores and background variables crosses this threshold he/she is offered admission. The recommended cut-off is chosen with a view to maximize the expected outcome or minimize the extent of misclassification – i.e., denying admission to high expected outcome types or offering admissions to low expected outcome types of applicants (c.f., Sawyer (2010)).

Despite the large literature on the first question cited above, there does not appear to be a corresponding systematic exploration of the second question – that of evaluating current admission policies– in educational statistics. The motivation for studying the second question is that widening access to good universities is generally considered to be a key mechanism for promoting social mobility. Therefore, under-representation of low SES communities among the entering students, especially at relatively elite colleges, is flagged as suggestive of unfair admission policies.[1] Colleges typically respond to such allegations by saying that they intend to admit students with the best academic potential, irrespective of their SES and such objectives can lead to uneven demographic compositions of entering classes. Testing whether this claim is correct is important from a policy perspective because the conclusion would inform what steps are necessary to promote fair access. For example, if admissions are indeed outcome-oriented, then good students from lower SES should be encouraged to apply in order to increase their presence in elite colleges; conversely, if admissions are, say, racially motivated (i.e., taste-based), then additional interventions may be needed in the admission process itself to promote fairness. This issue is substantively distinct from (i) the differential performance of different SES or demographic groups in aptitude tests (lower SES applicants perform worse on average), which is related to the issue of test-fairness and on which there exist several systematic studies, (ii) lower aptitude test scores proxying for lower SES in explaining post-enrolment outcomes (mixed evidence, c.f., Sackett et al (2009)) and (iii) the differential average college-outcomes of different demographic groups at the same value of aptitude test scores (black applicants have lower expected scores than Caucasian ones (Noble (2003)). These latter issues pertain to the average outcomes and are not directly relevant to testing academic fairness because academically fair – i.e., outcome oriented admissions – have implications for the outcomes of the marginal, but not the average, admitted candidates across demographic or SES categories. To see why, note that if the marginal male admit has a lower expected outcome than the marginal female admit then one can admit the "next best" female candidate instead of the marginal male admit and increase the expected overall outcome. So outcome-oriented admissions must equate the expected outcome across marginal

[1] See for instance, the recent newspaper stories: http://www.guardian.co.uk/education/2011/mar/20/oxford-universit,y-entrance-requirements-row,

or http://www.guardian.co.uk/education/2009/aug/19/oxford-university-men-places-women admitted candidates in every demographic/SES group. Violations of such inequality suggest that the college must place emphasis on non-academic attributes such as gender, race or SES of the applicant *beyond these attributes' impacts on his/her potential academic outcomes* and, in that sense, is "taste-based". On the other hand, the average outcome across all admitted males bears no relation to that across all admitted females, no matter whether admissions is outcome-oriented or not. This key but subtle distinction between the average and the marginal, to the best of our knowledge, is not recognized or used in the educational statistics literature and in public debates about unfair access to elite universities. But this distinction is quite well-recognized in studies of discrimination in economics (c.f., Heckman (1998), Persico (2009)). We intend to bring these insights from the economics literature to bear upon the question of testing academic fairness of admissions.

However, it is still non-trivial to identify the marginal admitted groups from admissions data that are typically available. The present paper will develop a method of identifying such marginal admits and then use them to develop a formal statistical test of whether current admission protocols are consistent with (academic) outcome maximization goals. The key insight is that under appropriate regularity conditions, the admission thresholds faced by every category of applicants can be identified from data on post-admission performance of just the admitted students; it is not necessary to identify the (unobserved) distribution of potential outcomes for non-admitted applicants. These thresholds, in turn, will equal the expected outcome of the marginal admit in every category and a test of outcome-oriented admissions can then be based on testing the equality of these thresholds across the categories. In the present paper, we develop the necessary econometric tools for (i) conducting a test of academically fair admissions and (ii) calculating the efficiency loss – in terms of academic achievement foregone – which would result from the use of unequal thresholds (say, due to affirmative action) across demographic subgroups.

2 Set-up and Methodology

Let W denote an applicant's covariates which are observed by the university with support $\mathcal{W}$ and let g denote the discrete component of W capturing the group identity of the applicant (such as sex or type of high school attended). For an applicant i, let Y_i denote his/her future academic performance if admitted to the university. Let $\mu(w)$ denote $E(Y|W=w)$. Let $c \in (0,1)$ denotes the fraction of applicants who can be admitted, given the number of available spaces.

Admission protocols: An admission protocol is defined by a probability $p(\cdot) : \mathcal{W} \mapsto [0,1]$ such that an applicant with characteristics w is admitted with probability $p(w)$. A generic protocol may be described as the solution to the problem:

$$\max_{p(\cdot)} \int h(w)\, p(w)\, y dP_{Y,W}(y,w), \text{ s.t. } \int p(w)\, dP_W(w) \leq c.$$

Here, $h(w)$ denotes a non-negative welfare weight, capturing how much the outcome of a w-type applicant is worth to the university. For affirmative action policies, $h(\cdot)$ will be larger for applicants from disadvantaged SES or under-represented demographic groups. The solution to the above problem takes the form (c.f., Bhattacharya (2011)):

$$p^*(w) = \begin{cases} 1 \text{ if } \beta(w) > \gamma, \\ q \text{ if } \beta(w) = \gamma, \\ 0 \text{ if } \beta(w) < \gamma, \end{cases}$$

where

$$\begin{aligned} \beta(w) &:= h(w)\, \mu(w), \\ \gamma &:= \inf\left\{a : \int 1\{\beta(w) > a\}\, dF_W(w) \leq c\right\}, \end{aligned}$$

and $q \in [0,1]$ satisfies

$$\int (\{1(\beta(w) > \gamma) + q1(\beta(w) = \gamma)\})\, dF_W(w) = c.$$

The result basically says that the planner should order individuals by their values of $\beta(W)$ and first admit those W whose $\beta(w)$ is the largest, then those for whom it is the next largest and so on till the budget is exhausted. If the distribution of $\beta(W)$ has point masses, then there could be a tie at the margin, which is then broken by randomization (hence the probability q). In the absence of any point masses, the optimal protocol is of a simple threshold-crossing form $p^*(w) = 1(\beta(W) \geq \gamma)$.

Academically fair admissions: We define an academically fair admission protocol as one which maximizes expected performance of the incoming cohort subject to the restriction on the number of vacant places. Such an objective is fair in the sense that the expected performance criterion gives equal weight to the *outcomes* of all applicants, regardless of their value of W, i.e., $h(w)$ is a constant. In this case, the previous solution takes the form $p^*(w) = 1\{\mu(w) \geq \delta\}$, where δ solves

$$c = \int_{w \in \boldsymbol{w}} 1(\mu(w) \geq \delta)\, dP_W(w).$$

Thus δ is the $(1-c)$th quantile in the marginal subjective distribution of the random variable $\mu(W)$. If $\mu(W)$ has a Lebesgue density that is not identically zero on any open interval contained in its support, then δ is uniquely defined.

It is important to note that maximizing the mean outcome gives equal weight to the outcomes of all applicants; consequently, the δ defined above does not depend on W. To see why, suppose one of the covariates in W is gender and assume that the admission threshold for women, δ_f is strictly larger than the threshold for men, δ_m. Then by admitting women whose expected outcome is just below δ_f but still higher than δ_m (and rejecting men who are at or above δ_m but lower than δ_f) the expected outcome of the entering cohort can be increased. So different thresholds cannot be consistent with the objective where the expected outcome of every individual is weighed equally in deciding whom to admit. The above argument presupposes the technical requirement that there are identifiable (by the university) women whose expected outcomes are "just below δ_f" – which is guaranteed by a positive density assumption below.

Let $W=(X,G)$ and let Ω_g, Υ_g denote respectively the support of X and $\mu(X,g)$, given $G=g$. We will assume that the random variable $\mu(X,g)$ has the same support for every g. For example, suppose G denotes gender and X contains one or more continuous variables like previous test scores. Then the above assumption basically says that given any value $x\in\Omega_{male}$, there exists an $x'\in\Omega_{female}$ such that $\mu(x,male)=\mu(x',female)$. Indeed, it seems very unlikely that there will be a range of possible values of the (conditional) expected outcome for men where no women's conditional expected outcome lies (or vice versa). Note also that we do not require Ω_g to be identical across g.

Under the above assumption, for every fixed g, there will exist $x(g)\in\Omega_g$ such that $\mu(x(g),g)=\delta$; so we can define individuals in subgroup g with $X=x(g)$ to be the marginal admits in subgroup g. Our identification strategy for δ is based on calculating $\mu(x(g),g)$, i.e., the expected outcome of the marginal admit for various values of g. Then testing academic fairness of admissions amounts to checking if the δ's are identical across g.

We now formally set up the econometric model on which we will base our test of fairness. To keep the exposition focused, we will replace G by gender with the understanding that we can replace G by any other characteristic.

Econometric model and assumptions: Let Y denote the relevant college outcome, such as final year GPA and let D denote the dummy variable indicating whether an applicant is offered admission and A denotes a dummy for whether the offer was accepted.[2] Our

[2]Typically, not all admission offers are accepted. This can happen in two ways – one, some candidates can

econometric model is then given by the following equations:

$$\begin{aligned} D_i &= \mathbf{1}\left\{\mu^*\left(X_i, G_i\right) \geq \gamma\left(G_i\right)\right\}, \\ \mu^*\left(x, g\right) &= E\left(Y|A=1, X=x, G=g\right)+\varepsilon, \end{aligned}$$

where G_i denotes each individual's gender, i.e., $G_i = f$ or m (female or male); (X_i, G_i, D_i) is observed for any individual $(i = 1, \ldots, n)$. The outcome Y_i is observed only when $A_i = 1$. The quantity $\mu^*(x, g)$ represents the expected outcome for an (x, g) type applicant, as perceived by the admission officer handling applicant i's file. This subjective expectation is assumed to have formed from the officer's experience about the performance of previously admitted (x, g) type candidates who had enrolled and for whom the college outcomes were observed. The error ε represents the deviation of an individual tutor's subjective expectation from the true mathematical expectation.

We are interested in estimating the coefficients $\gamma(g) = \gamma_g$ for $g = female$ (f) and $g = male$ (m). The difference $\gamma_f - \gamma_m$ represents a measure of deviation from outcome based admissions and the extent to which the admission process is "biased" against female applicants, from the academic efficiency point-of-view.

For the identification of γ_g, we impose the following conditions:

Assumptions: (i) $\{(X_i, G_i, D_i, Y_i A_i)\}_{i=1}^n$ is an i.i.d. sequence; (ii) ε has strictly positive Lebesgue density around 0, given X, G and $\text{med}(\varepsilon \mid X, G) = 0$.

Discussion: The error term ε_i incorporates the error in subjective expectation of the specific admission officer handling applicant i's case. This allows for admission to be non-deterministic, given X, G. In other words, the admission cut-off (for expected performance) faced by applicants of type g is $\gamma_g + \varepsilon$ which varies randomly across g-type applicants with a median value of γ_g. For academically fair admissions, as explained above, one would expect γ_g to be identical across g. The key implication of the two equations above is that

$$D = \mathbf{1}\left\{\mu_1\left(X, G\right) > \gamma_G + \varepsilon\right\}, \tag{1}$$

where $\mu_1(x, g) = E(Y|A = 1, X = x, G = g)$, which is identifiable from outcome data on past enrolled students.

receive and accept offers from other universities and two, some candidates may fail to satisfy the conditions specified in the original offer, such as securing a certain percentage in their school-leaving examinations. Such nonrandom acceptance is not problematic if, as is quite natural, tutors form their subjective expectations, based on the performance of students who had actually entered because those are the only performances they actually observe.

Identification: To see how γ_g is identified, note that when $\Pr(D=1|X,G=g)=1/2$, we must have that $\Pr\left[\mu_1(X,g)-\gamma_g\geq\varepsilon|X,G=g\right]=1/2$, so that, by assumption (ii),

$$\mu_1(X,g)-\gamma_g=0.$$

Therefore, it holds that

$$E\left[\mu_1(X,G)-\gamma_G \mid \Pr(D=1|X,g)=1/2, G=g\right]=0. \tag{2}$$

This suggests that for group g, the threshold γ_g can be identified as the average performance of those g-type candidates who had enrolled in past years and whose x's imply a predicted probability of receiving an offer equal to 0.5. Our formal result on identification of γ_g is facilitated by the following assumption:

Assumption: (iii) Conditional on each possible value g of G, the random variable $\mu_1(X,g)-\gamma_g$ has a strictly positive Lebesgue density on an open interval around 0.

Observe that assumption part (ii) lets us identify the marginal admits as those x for which the probability of admission is close to a half. Part (iii) guarantees that for every fixed g, the set $\left\{x\in\Omega_g:\mu_1(x(g),g)=\gamma_g\right\}$ is nonempty and this is the limiting set over which $\mu_1(\cdot,g)$ will be averaged to yield γ_g. Finally, $\mu_1(x,g)$ is directly identified from the data on previously admitted students.

Estimation: If X contains continuous components, then typically the estimated probability of getting an offer will never be exactly $1/2$; so we consider those (x,g)'s for which the predicted probability is close to a half. And the extent of closeness is guided by a bandwidth h. Formally, we estimate γ_g by:

$$\hat{\gamma}_g=\frac{\sum_{i=1}^{n}K_h\left(\hat{p}(X_i,G_i)-1/2\right)\mu_1(X_i,G_i)\mathbf{1}\{G_i=g\}}{\sum_{l=1}^{n}K_h\left(\hat{p}(X_l,G_l)-1/2\right)\mathbf{1}\{G_l=g\}}. \tag{3}$$

where $K_h(z)=K(z/h)/h$; $K(\cdot)$ is a kernel function; h is a smoothing parameter (bandwidth); $\hat{p}(x,g)$ and $\mu_1(x,g)$ are first-step estimators of $p(x,g)$ ($:=\Pr[D_i=1|(X_i,G_i)=(x,g)]$) and $\mu_1(x,g)$, respectively. This is exactly the average predicted outcome for those (x,g) types for whom the predicted probability of getting an offer, i.e., $p(x,g)$ is close to a half.

Below, we establish formally (proof in appendix) that our estimator $\hat{\gamma}_g$ has an asymptotic distribution, given by

$$\sqrt{nh}\left[\hat{\gamma}_g-\gamma_g-h^2\mathbf{B}(g)\right]\xrightarrow{d}N(0,V(g)),$$

where

$$\mathbf{B}(g) \; : \; = \int_{\mathbb{R}} u^2 K(u)\, du \left[\frac{(\partial/\partial z)\, m(z,g) \times (\partial/\partial z)\, \nu(z,g)}{\nu(z,g)} + \frac{1}{2}\left(\partial^2/\partial z^2\right) m(z,g) \right] \Bigg|_{z=1/2};$$

$$\mathbf{V}(g) \; : \; = \int_{\mathbb{R}} K^2(u)\, du \frac{\mathrm{Var}\left[\mu(X_i,g)\, | p(X_i,g) = z\right]}{\nu(z,g)} \Bigg|_{z=1/2}.$$

$\nu(z,g)$ is the probability function of the random variables $p(W_i)$ and

$$m(z,g) := E\left[\mu(W_i) \;\; | \; p(W_i) = z, \; G_i = g\right] = E\left[\mu(X_i,g) \;\; | \; p(X_i,g) = z\right].$$

Implementation:

- For the second-step bandwidth h, the optimal rate (minimizing the MSE) is

$$h = O\left(n^{-1/5}\right).$$

  However, given this rate, we incur the asymptotic bias and the CI based on the limit normal distribution in Lemma **??** would be technically invalid unless we estimate $\mathbf{B}(g)$. However, it is not easy to estimate $\mathbf{B}(g)$, since it involving the derivatives of relevant functions. So, we

  - set $h = o\left(n^{-1/5}\right)$, say, $h = O\left(n^{-1/5}/\log n\right)$; or
  - use $h = O\left(n^{-1/5}\right)$ but still use the normal approximation based on Lemma **??** (in this case, we obtain something called "confidence band" or "variable band," which is not a CI).

3 Welfare calculations

We now develop a method of inferring the efficiency costs of progressive admissions policy. To see how to do this, suppose that progressive motives lead a university to choose a lower threshold for female candidates, i.e., $\gamma_f < \gamma_m$. Then, one can evaluate the net cost, in terms of academic efficiency, of this policy as the expected outcomes of all male candidates who were not admitted at the higher threshold but would have been admitted had their threshold been lowered to γ_f. We will develop a theory of inference for this parameter, formally defined as:

$$\omega_1 = E\left[\mu_1(X,m) \times 1\left(\gamma_f \le \mu_1(X,m) \le \gamma_m\right) | male = 1\right] \times \pi_m,$$

where π_m denotes fraction male. Alternatively, relative to a common intermediate threshold $\gamma_f < \gamma < \gamma_m$, the net welfare loss of the current practice may be computed as the "displacement cost" of replacing high ability males with low ability females:

$$
\begin{aligned}
& \omega_2(\gamma) \\
= \; & E\left[\mu_1(X,m) \times 1\left(\gamma \leq \mu_1(X,m) \leq \gamma_m\right) | male = 1\right] \times \pi_m \\
& -E\left[\mu_1(X,f) \times 1\left(\gamma_f \leq \mu_1(X,f) \leq \gamma\right) | male = 0\right] \times (1-\pi_m).
\end{aligned}
$$

The intermediate γ can be chosen so that the aggregate number of admits, if the common γ is used, will remain identical to current number of admits...

4 Empirical Application

We first present a brief description of the data. Our current data cover 2 cohorts of Oxford PPE students who started their study in 2008 and 2009 respectively. We focus on those applicants who had taken the TSA (an aptitude test comparable with the SAT in the US), and the GCSE – a standardized school-leaving examination in the UK. The following tables summarize for each gender the key variables. Of these, col1inter is the interview score at admissions, got_in is a dummy for whether the candidate was offered admission, independent_school is a dummy for whether the applicant was educated at a fee-paying (i.e., private) school. The variable prelim_tot is the overall score after the first year examinations which are obviously available for admitted candidates only.The above table shows that many more males apply to read PPE, males applicants have better TSA distributions, worse interview scores and male admits score an average of 4 points (out of 300) higher in the first year exams. These differences are statistically significant at 5%. There is no significant difference in acceptance rates between male and female candidates.

We now test if the marginal admitted male and the marginal admitted female student have identical expected first year scores. To do this, for each gender, we compute the expected score as a linear function of age, GCSE score, overall TSA, interview score and whether the applicant came from an independent school. Using the 0 median restriction on expectation errors, as explained above, we calculate the threshold faced by each gender as the average of expected first year scores for admitted applicants whose probability of being admitted is predicted (through a probit) to be close to 0.5. In the following table, we show the estimated difference together with the standard errors for a range of bandwidths (which define "closeness to 0.5") and an Epanechnikov kernel.This table strongly suggests that male

		Female			
Variable	Obs	Mean	Std. Dev.	Min	Max
age	609	17.99	1.24	16.25	36.95
gcsescore	609	3.82	0.26	2.38	4.00
tsaoverall	609	61.96	7.72	39.70	101.90
col1inter	609	62.46	11.03	5.67	87.00
independent_school	609	0.34	0.47	0.00	1.00
got_in	609	0.28	0.45	0.00	1.00
prelim_tot	150	183.71	15.76	137.00	217.00
		Male			
Variable	Obs	Mean	Std. Dev.	Min	Max
age	923	18.12	1.62	16.21	54.08
gcsescore	923	3.75	0.35	1.00	4.00
tsaoverall	923	65.46	7.94	35.60	101.90
col1inter	923	63.46	11.44	0.00	95.00
independent_school	923	0.39	0.49	0.00	1.00
got_in	923	0.31	0.46	0.00	1.00
prelim_tot	247	187.62	19.05	106.00	233.00

Figure 1:

bw	male thld	std dev	female thld	std dev	male-fem	std. Err	z
0.08	187.80	3.73	179.78	3.07	8.02	4.83	1.66
0.80	188.58	2.47	181.70	2.59	6.88	3.58	1.92
8.00	186.88	2.37	180.24	2.43	6.64	3.39	1.96

Figure 2:

bw	Indep thld	std dev	state thld	std dev	indep-state	std err	z
0.09	185.19	4.29	190.08	5.76	-4.89	7.18	-0.68
0.90	182.06	3.32	187.88	1.71	-5.82	3.73	-1.56
9.00	183.13	2.23	185.02	2.51	-1.89	3.36	-0.56

Figure 3:

applicants face a threshold that is higher than that faced by female applicants by about 6-8 points out of 300. Therefore, although male and female acceptance rates are nearly equal, male candidates have to cross a higher barrier of expected performance than females in order to get admitted.

We then repeat the analysis reversing the roles of gender and school background, i.e., we use gender as an explanatory variable and test if applicants from independent school face a higher threshold than their counterparts who apply from state-funded schools.Now, we see that there is hardly any difference in the thresholds faced by the two groups. Although the estimated thresholds for private school applicants are lower, the difference is never statistically significant at 5%....

5 Bibliography (incomplete...)

Becker, Gary (1957). The economics of discrimination, University of Chicago Press.

Bhattacharya, D. (2009). Inferring Optimal Peer Assignment from Experimental Data. Journal of the American Statistical Association, Jun 2009, Vol. 104, No. 486: 486–500.

Bhattacharya, D. & Dupas, P. (2010). Inferring Efficient Treatment Assignment under Budget Constraints", NBER working paper number 14447, under revision at Journal of Econometrics.

Heckman, J. (1998). Detecting discrimination, Journal of Economic Perspectives-Volume 12, Number 2, 101-116.

Kobrin et al (2008). Validity of the SAT for Predicting First-year College Grade Point Average, College Board, New York.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. Psychological Bulletin, 127, 162-181.

Lindskog, F., McNeil, A., & Schmock, U. (2004). Kendalls tau for elliptical distributions.

Millsap, R. & W. Meredith (2007). Factorial Invariance: Historical Trends and New Developments, in Factor Analysis at 100, ed. Robert Cudeck and R. MacCallum, Lawrence Earlbaum and Associates, New Jersey.

Parks, G. (2011). Academic Performance of International baccalaureate students at Cambridge, by school, http://www.cam.ac.uk/admissions/undergraduate/research/docs/ib_performance.pc

Persico, N (2009). Racial Profiling? Detecting Bias Using Statistical Evidence. Annual Review of Economics, volume 1.

Sawyer, R. (1996). Decision Theory Models for Validating Course Placement Tests. Journal of Educational Measurement, Vol. 33, Issue 3, 271-290.

Sawyer, R. (2010). Usefulness of High School Average and ACT Scores in Making College Admission Decisions, available online at: http://www.act.org/research/researchers/reports/pdf/ACT_ 2.pdf

Sackett, P., Kuncel, N., Arneson, J., Cooper, G., Waters, S. (2009). Socioeconomic Status and the Relationship Between the SAT and Freshman GPA - An Analysis of Data from 41 Colleges and Universities, available online at: http://professionals.collegeboard.com/data-reports-research/cb/SES-SAT-FreshmanGPA