How We Justify and Perpetuate the Wealthy, White, Male Academic Status Quo Through the Use of Biased Admissions Requirements

By:

Theodore Micceri, Ph.D.

Paper Presented at the Florida Association for Institutional Research Annual Conference, Cocoa Beach, FL. Feb 25-27, 2009

How We Justify and Perpetuate the Wealthy, White, Male Academic Status Quo Through the Use of Biased College Admissions Requirements

This paper demonstrates empirically how a trend that began during World War II helps ensure that males and society's upper class of wealth and power (usually white) gain unjustified advantages when it comes to enrolling in prestigious public universities as a direct result of standardized admissions tests. Higher education has historically been, at least until the proletariat revolutions of the 20th Century, almost the exclusive territory of the aristocracy's young males. This was a place they could send their young to avoid the danger of interacting with city workers.[1] As an example, between 1890 and 1900, fewer than five percent of Americans aged 18 to 21 attended higher education (Goldin & Katz, 1999, p. 41). By 1970, five percent had risen to roughly 70 percent, meaning that large numbers of "commoners" were rubbing elbows with the elite, partly due to the 1965 Higher Education Act which stated that colleges couldn't turn away applicants merely because their families were poor. Such interaction with society's riffraff had traditionally been avoided by the aristocracy through the use of exclusive (expensive) private colleges. In theory, public universities, which are paid for by the taxes of all, should admit quality lower class students. In the latter part of the 20th Century and the early years of the 21st Century, as the United States experiences increasing wealth disparity (Sahadi, 2006; Witte & Henderson, 2004), we see a disturbing tendency to exclude "commoners". Haycock & Gerald (2006, p. 3) state, regarding public flagship institutions: "Even as the number of low-income and minority high school graduates in their states grows, often by leaps and bounds, these institutions are becoming disproportionately whiter and richer." Others echo such statements (Heller & Marin, 2002; Astin & Oseguera, 2004; Sacks, 2007). The tool of choice for discrimination is standardized tests justified by the phrase "stringent admissions requirements," see Astin & Oseguera (2004). In the Florida State University System (SUS) more stringent admissions recently resulted from a 70% increase in First Time in College (FTIC) matriculations during a time when constant dollar state funding only increased by 11 percent[2] (1996 to 2003) which gave institutions the choice of restricting access or restricting service. Apparently SUS institutions generally chose to restrict access more than service.

Although the initial hope was that standardized tests would reduce class bias, in reality, they increase the class bias effect, as data from over one million Florida State University System (SUS) applicants in this study demonstrates. Two factors stimulated the current work (1) earlier research conducted in an attempt to understand the recent movement of underrepresented minorities from direct entry into the SUS to transfer from community colleges, and (2) Gibson's (2001, p 1.) claim that: "…The SAT measures, above all else, class, gender, and race…" This study addresses the research question: Do any consistent score differences occur on standardized tests between different sexes or race/ethnicities (a proxy for affluence) for students exhibiting the same academic performance levels as measured by High School Grade Point Average?

Background

Between 1996 and 2003, the Florida SUS experienced a 70 percent increase in the number of First Time In College (FTIC) students and a 33 percent increase in total new undergraduate enrollees during a time when HEPI-based constant dollar funding increased by 11 percent (SHEEO, 1997; Palmer, 2004; Commonfund, 2006). Thus, SUS institutions were faced

[1] Thus, most major state public universities locate in small, comparatively affluent towns like Albany, Austin, and Athens rather than New York City, Dallas or Atlanta.

[2] Using the Higher Education Price Index (HEPI), to provide a more realistic constant dollar estimate than that of the highly inaccurate Consumer Price Index (CPI).

with the choice of either reducing service or reducing access and they apparently chose the latter. As a result, it has become increasing difficult for high school students to gain admission to an SUS institution (Vogel, 2006), and this same effect has occurred nationally (Astin & Oseguera, 2004). During this time, despite increasing percentages of minority high school graduates, who increased nationally from about 37 percent to 42 percent of the population (CDC, 2006), their matriculation directly from high school into the more prestigious SUS 4-year institutions remained relatively flat, growing by only 1.6 percentage points (36.4% to 38.0%). During the same period, minority entry to the SUS among Community College Transfers (CCT) increased by 9.4 percentage points (27.8% to 37.2%). While African American representation among SUS FTIC students remained identical between 1996 and 2003 (18.3% and 18.4%), they showed a 30 percent representation increase among CCT students (9.8% to 12.7%). Students classified as "Other" more than doubled their CCT representation, from 2.9 percent to 6.2 percent,[3] while Hispanic students increased from 15.1 percent to 18.3 percent. Females remained more stable in both populations, but showed a 2.5 times greater increase as CCT students (2.4%) than FTIC students (0.9%).

The increase in minority and female percentages among CCT transfers elicited an investigation of possible biases in admissions that might adversely influence minority and female direct entry to SUS institutions. An analysis of 164,378 SUS fall FTIC applicants (2001 to 2003) comparing mean SAT scores between whites and minorities and between sexes at each 10^{th} of a GPA point showed a bias favoring whites and males. At the same GPA scale point (e.g. 3.6), the mean difference between whites and minorities averaged 62 points across years. Males exhibited an even greater average advantage of 75 points over females across years (Borman, Workman, Miller & Micceri, 2006). On the topic of differences between individuals ETS (2001) states: "The user can be reasonably confident that a score difference of around 60 points or more indicates a 'true' difference in ability between two test takers." One would assume that this should be more true for comparatively stable group means than for somewhat unreliable individual scores. Therefore, if a "true difference" between males and females actually exists as these test score differences suggest, one must ask: a "true" difference on what, because females outperform males everywhere in education, even in Engineering (Micceri, 2005). These background and findings prompted the current, broader and more intensive study of bias in standardized tests.

The History of Standardized Tests as Admissions Criteria

Understanding how standardized tests came to replace the traditional essay as a primary selection tool for college admission helps explain much of what has happened over the past 65 years. The initial push to use IQ tests as criteria for officer selection in the military was led by Terman and Yerkes, executives in the American Eugenics Society (AEC), who wanted to purify the race of low-grade and degenerate groups such as minorities and the poor (Gibson, 2001). From an opposite, egalitarian perspective, James Bryant Conant, Harvard's president, believed that in the half century leading up to 1940, the U.S. had gone from being a classless society to one that was falling under the control of a hereditary aristocracy. Conant hoped to use the SAT to select students for college who had virtue and talent, assuming that the two went hand-in-hand. Regarding the SAT he wrote: "...we have before us a new type of social instrument whose proper use may be the means of salvation of the classlessness of the nation..." Conant sought to use this new social instrument (the SAT) to sort and slot the entire populace on the basis of their test-

[3] Approximately half of "other" students are resident aliens, and this group also includes multi-racial students.

defined intelligence in the name of creating a perfected, classless and democratic America modeled on Plato's Republic (Lemann, 1999, 1999a).

Wars provide opportunity for restructuring society and Conant, realizing this, moved quickly to establish the current testing regime after WW II began. Just after Pearl Harbor, he replaced the old essay tests for college admission with the SAT for all applicants at Harvard. In 1943, a revised SAT was administered to more than 300,000 people nationwide for officer-selection purposes (Lemann, 1999a).

Immediately after the war, Conant, through an adept series of bureaucratic maneuvers, arranged for all the leading education tests and testing organizations in the country to be merged into a new, private, non-profit entity that would effectively hold a monopoly in the field, the Educational Testing Service (ETS). Because the service is private, it does not answer to the populace, as do government-run testing organizations in most countries. Because these aptitude tests were free to colleges, represented far less work on the colleges' part than traditional essays, and were thought to be objective and valid, following nationwide marketing by ETS, they spread rapidly as a college selection criterion. Conant's purpose was to create an objective measure that could identify his aristocracy of virtue and talent even when located in poorer communities (Lemann, 1999a, p. 53). Unfortunately, these tests have not proven objective and precisely the opposite of what Conant hoped has turned out to be true due to a consistent bias favoring affluence as numerous recent articles attest, for example Heller & Marin, 2002, Astin & Oseguera, 2004; and Sacks, 2007. This effect was perhaps most clearly elucidated by renowned demographer Harold Hodgkinson (Hodgkinson, 1999, p. 17): "SATs predict one thing beautifully, but it's not the grades students will earn as freshmen; it's the household income of the test-takers. For every $10,000 increase in household income, math and verbal scores go up a minimum of nine points." One might ask, "How can this occur in scientifically developed measures like standardized tests?" As the literature review below shows, partly it results from very non-scientific political and cultural influences in test development, and partly from a specific set of biases that consistently favor the affluent.

The Value of Standardized Tests as College Admissions Criteria

The following statement summarizes the results of thousands of historic studies conducted at almost every college and university on the relationship between prior academic variables (GPA and tests) and college performance: Relationships between tests and any performance measures in college becomes essentially zero when one controls for primary predictive factors such as prior performance (GPA), affluence (which may be imprecisely estimated by part-time/full-time enrollment) and gender.

Even test makers themselves do not claim that standardized tests measure either achievement or school outcomes (Bracey, 1997), although they do correctly claim a low-level relationship with first semester college grades (Murphy, 2000; Elert, 1992). This statement is supported by many studies which report positive simple relationships (correlations) between tests (ACT, SAT, GRE or GMAT) and first semester grades in college, that generally range between ***r*** = .15 and ***r*** = .25. A few related studies include: Adelman, 2006; Astin, 1993; Bangura, 1995; Bieker, 1996; Chernyshenko & Ones, 1999; Florida SUS, 1995; House, 1998; House, Gupta & Xiao, 1997; Lindle & Reinhart, 1998; Morrison & Morrison, 1995; MSCHE, 1997; Onasch, 1994; Paszczyk, 1994; Sternberg & Williams, 1997; Stricker, & Rock, 1993, Waugh, Micceri & Takalkar, 1994; Xiao, 1998.

This relationship with first semester grades and high school grades probably results largely from method variance. When you use the same method in measurement of different targets (e.g. multiple choice tests), a proportion of the variance attributes to the method. This, rather than G, is mostly likely a primary reason that multiple-choice tests of different subjects tend to show relatively strong correlations with each other, but with almost nothing else. Regarding the value of predicting first semester grades, Alexander Astin (cited in Elert, 1992) notes: "In a very practical sense, the student's ability to stay in college is a more appropriate measure of his success than is his freshman GPA."

Crouse & Trusheim (1988), using data from the National Longitudinal Study (NLS) of the high school class of 1972, and four different measures of undergraduate success, calculated that using the SAT in admissions adds between 0.1 and 2.7 additional correct forecasts per 100 applicants. Some will argue that restriction of range reduces the relationship between test scores and college performance, and this is certainly somewhat true. However, most of the studies cited were conducted at less restrictive institutions, where almost the full range of possible test scores frequently occurs. Further, the current study is not influenced by this effect, because it does investigate relationships, but only the test scores that associate with specific high school GPA values.

Among the thousands of studies conducted on this topic, (Breland & Minsky, pp. 149-153) in their summary, report that the strongest relationship between SAT and grades occurred where the 1978 SAT-Verbal matched random chance at a New Jersey college 59 percent of the time (***r*** = .41), and weakest was the 1972 SAT-Verbal at an Indiana school, which was essentially random (99.96% of the time). To give some idea of how extensive such research has been, among the highest relationships in the testing literature between the SAT and anything other than another test is a correlation of ***r*** = .66 reported by French & John (1967) between SAT scores and uric acid levels in the blood, a stress indicator. Such studies show just how widely ETS and ACT have searched, with almost no success, for something their products predict empirically.

Why Standardized Tests Generally Lack Validity

The preceding researches suggest that standardized tests add little to prediction. The following story shows one major reason for this phenomenon:

> Forest Gump died and is trying to get into heaven. At the pearly gates, St. Peter says that he must pass a test to enter the Kingdom of Heaven, so Forest says: "OK." St. Peter says: "You must correctly answer three questions:
>
> 1. What is God's name?
> 2. How many seconds are there in a year?[4]
> 3. How many days starting with the letter 'T' are there in a week?"
>
> So, Forest goes off to think, and when he comes back, St. Peter asks how many days in the week start with the letter 'T.' Forest replies, "That's easy, two: Today and Tomorrow." St. Peter says, "Well, that isn't exactly what I thought I

[4] This is a really tough question, because not only are there 60 seconds in a minute, 60 minutes in an hour, 24 hours in a day and 365 days in a year, but to get it correctly, one must add ¼ day each four years, and ¼ day each four hundred years and I believe there are more adjustments necessary when one uses a less precise Calendar, such as the Julian, instead of a more readily computed ones such as the lunar calendar used in many African and Muslin countries.

was asking, but you are correct. So, how many seconds are there in a year?" And Forest says, "There are 12." St. Peter looks startled and asks, "How did you figure that?" Forest replies, "Well, there's January 2nd, February 2nd, and so on." St. Peter says, "I see that you are correct, although again, that wasn't what I thought I was asking. OK, so what is God's name?" To this, Forest replies, "Ah shucks, that's easy, we learned that in Sunday school. God's name is Andy." St. Peter looks startled again and says "Andy?" Forest says, sure, it starts off that song, 'Andy walks with me, Andy talks with me...'"

This story clearly points out one primary and vital flaw in standardized language-based tests: the utterly fallacious assumption that only one legitimate interpretation exists for a sentence/question/answer. This produces a strong and consistent bias against second language learners or anyone coming from a different language background than Standard English (Underbakke, 2000). Such individuals frequently view language from a different perspective than that of the typical middle-to-upper-class American test writers and takers. As the father of a son born in Peru, who's native language is Spanish, I have the opportunity to see this regularly. When Juan is trying to answer written questions, he frequently sees two and sometimes as many as three different legitimate interpretations for various homework questions. Until he points them out and explains them, I usually would not have thought about the question in that way, but after his explanations, I can see he is correct, just like Forest Gump.

Biases Common to All Forms of Standardized Tests

Although many mistakenly think them objective, standardized tests exhibit several biases against certain subpopulations of test-takers. As a result, tests consistently underestimate these groups' knowledge and performance. In measurement, biases can arise in some interesting ways. For example, when researching his non-language IQ tests, Raven (1939) discovered that the physical location of correct and erroneous answers consistently influenced (biased) student errors. The following are a few of the most obvious and common types of bias inherent to all standardized tests.

1. Students from households using non-standard English face a consistent disadvantage. No matter what their racial/ethnic group, poorer people in the United States tend to grow up in such environments. As an example of how such biases can lead to totally erroneous conclusions, early in the 20th century, when widespread testing began, immigrants from Poland and Italy were considered stupid and inferior because they scored poorly on IQ tests in their second language, English.

2. Regarding cultural biases, test creators tend to grow up in middle-class to upper class American households. The test items such individuals write tend to reflect their personal and cultural experiences. Those from poorer households, different countries, cultures or religions frequently lack experience with the topics discussed on tests. This type of bias first became apparent in the 1890's when researchers suggested the use of field trips to the country for city public school kids because city kids scored lower than country kids on standardized tests. Most of the test makers at that time had grown up on farms, and city kids knew little about animals or farming environments, which were common topics on tests created by former farm kids. As a friend from Australia once said: "If an average American were to take a test of required skills that was created by a Bushman, they would test at the moron level."

3. The great majority of high-stakes and standardized tests are speeded. This tends to work against those who are perfectionists, those who process information slowly, and those who are less likely to "take risks" (probably more generally true of females than males). Many people will not put down an answer until they are positive it is correct (perfectionists). Obviously, in a speeded test, such individuals have trouble answering enough questions to attain a high score. Additionally, some highly intelligent individuals process information far more slowly than do others (e.g. Albert Einstein, see below). In most situations this is not very relevant. However, when taking a timed test, it negatively biases estimates of knowledge, intellect and talent.

4. Large numbers of students today suffer from test anxiety. This begins to show in about the third grade and may cause a student to perform poorly on tests. Because prior performance influences test anxiety, students who are subject to items #1, #2, #3 and/or #6 will tend to perform less well and therefore be subject to greater test anxiety, which tends to further reduce their scores.

5. Standardized tests and almost all tests of any type tend to heavily reward short-term memory skills. Although this is a useful skill to have, it means that almost all such tests are biased against those who lack strong short-term memory skills no matter at what level their long-term memory skills mayThoTho be.

6. A fairly large proportion of today's student population has what is termed a learning disability. This may represent any of a number of different ways of looking at the world, but in most cases, these different perspectives or methods of processing information, associate with poorer performance in school and almost always with erratic test performance, which is used as an indicator of a learning disability.

7. As a rule, tests reflect a very limited perspective on a single type of intellectual process that may be termed abstract, logical/analytical, and which rarely requires any higher-order thinking from a respondent. Performance in school is also typically evaluated using tests of a similar nature. As a result, it is not uncommon for those who think differently, and who may be the greatest geniuses among us, to not exhibit the specific type of intelligence measured by either teacher's instruments or standardized tests (This bias frequently relates to #6, the learning ability/disability issue.). Robbins (1987, p. 40) notes some well-known examples from history:

> Albert Einstein's parents were sure he was retarded because he spoke haltingly until the age of nine, and even after that would respond to questions only after a long period of deliberation. He performed so badly in his high school courses, except mathematics, that a teacher told him to drop out, saying, "You will never amount to anything Einstein." Charles Darwin did so poorly in school that his father told him, "You will be a disgrace to yourself and all your family." Thomas Edison was called "dunce" by his father, "addled" by his high school teacher and was told by his headmaster that he "would never make a success of anything." Henry Ford barely made it through school with the minimum grasp of reading and writing. Sir Isaac Newton was so poor in school that he was allowed to continue only because he was a complete flop at running the family farm. Pablo Picasso was pulled out of school at the age of ten because he was doing so badly. His father hired a tutor to prepare him to go back to school, but the tutor gave up on the hopeless pupil. Giacomo Puccini, the Italian opera

composer, was so poor at everything as a child, including music that his first music teacher gave up in despair, concluding that the boy had no talent.

Most of the forms of bias noted above tend to exhibit a greater influence on less affluent groups, for several reasons. Aside from cultural differences relative to test creators, minority populations tend to fall into the "less affluent" category. For example, 2005 median incomes for males were: Other, $27,041, Hispanics, $27,380, African Americans, $34,433, whites, $46,807 and Asians, $48,693 (Webster & Bishaw, 2006). Many minority individuals, particularly those in the lower socio-economic classes, rarely experience Standard English in their homes or communities, and frequently not even in the low-performing schools they attend (note that the definition "low-performing" always results from language biased test performance).

FairTest (2006) documents a rather insidious and purposeful use of bias #2 above regarding scores on the SAT: "The SAT is designed solely to predict students' first year college grades. Yet, despite the fact that girls earn higher grades throughout both high school and college, they consistently receive lower scores on the exam than do their male counterparts."

The article reports that the gender gap favoring boys persists across all other demographic characteristics, including family income, parental education, grade point average, course work, rank in class, size of high school, size of city, etc. A study by Phyllis Rosser (1989), *The SAT Gender*[i] *Gap: Identifying the Causes*, found that the vast majority of questions exhibiting large gender differences in correct answer rates are biased in favor of males, despite females' superior academic performance. Rosser found that girls generally did better on questions about relationships, aesthetics and the humanities, while boys did better on questions about sports, the physical sciences and business. On this topic, Carol Dwyer, in her 1976 report notes that it is common knowledge among test-makers that gender differences can be manipulated by simply selecting different test items. Dwyer cites as an example the fact that, for the first several years the SAT was offered, boys scored higher than girls on the Math section but girls achieved higher scores on the Verbal section. ETS policy-makers determined that the Verbal test needed to be "balanced" more in favor of boys, and added questions pertaining to politics, business and sports to the Verbal portion. Since that time, boys have outscored girls on both the Math and Verbal sections. Dwyer notes that no similar effort has been made to "balance" the Math section, and concludes that "It could be done, but it has not been, and this suggests that either a conscious or unconscious form of sexism underlies this pattern (FairTest, 2006)."

Such information calls into question both the motives and trustworthiness of test developers, to whom we, as a nation annually entrust hundreds of millions of public dollars in addition to millions of hours of students time and, in fact, the very future of our youth.

In view of the preceding, it is not surprising that certain groups, particularly females, the less affluent, culturally distinct, and second language learners, score consistently lower than other groups. Thus, historical evidence suggests that these measures are not objective estimates of talent or knowledge.

Methods

This study addressed the following research question:

Do any consistent score differences occur on Standardized Tests between different sexes or race/ethnicities who exhibit the same historic academic performance as measured by high school GPA?

To assure adequate sample sizes in each cell, data from the Florida SUS Master Admissions files for all SUS FTIC applicants for the Academic Years (AY) 1997-98 through 2005-06 were submitted to analysis.

Variables and Data Analysis

High School GPA (GPA) - values are those reported in Admissions Files by SUS institutions. Students in high school obtain extra GPA points for taking AP, IB and Honors courses, thus the possible range of GPA values is from 0.0 to 5.0. For these analyses, in order to assure adequate sample sizes in each cell, the range was limited to 2.5 through 4.5 because some groups become comparatively rare at certain GPA levels (for example, Asians below 2.5 and most racial/ethnic groups above 4.5.)

Race/Ethnicity – Self-reported classification obtained from SUS institutions. Analyses were limited to Asian, African American, Hispanic and white, non-Hispanic. Similar results as those reported also occurred among students classified as Other, defined, as all others not classified as unknown. However, for brevities sake those are not reported here. Those classified as unknown were excluded from all analyses.

Gender – Self-reported classification obtained from SUS institutions. Students were classified as male or female. All unknown cases were excluded from analysis.

Bias – Defined as consistent error in one direction.

For brevities sake, the current paper reports only SAT Total Scores, although SAT Quantitative and Verbal subscores as well as ACT Composite Scores were submitted to analysis with almost exactly the same findings.

The Florida SUS Admissions Files – The Florida Master Admissions Databases are housed at the Northwest Regional Data Center in Tallahassee, Florida, and are produced by the Florida Department of Education. The Florida SUS Master Admissions Files provides relevant academic and demographic data for all applicants to any SUS institution.

GPA values were rounded to the nearest 10^{th} of a point and mean test scores were computed for each race/ethnic subgroup and for both sexes separately for all students within each group at each GPA level. SAS version 9.1 was used to compute the subgroup score values, and between group differences were obtained using Microsoft Excel 2004. Charts were created using Microsoft Excel.

Limitations

It is possible that high school GPA values for the same performance level differ from one school or school district to another. The possible detrimental influence of such effects on findings was addressed by assuring large samples for each cell (The smallest cell size was 149 cases.). Large samples from multiple schools and years should weaken any such biasing effects.

One might ask the question of whether there should be a relationship between GPA and tests. This is the only claim regarding the validity of these tests for college selection that the test makers themselves assert (Fairtest, 2006). Further, as Figures 1 and 2 show, a monotonic relationship exists between high school GPA and test score values that remains invariable across all race/ethnicity and gender subgroups.

Results and Discussion

Sample

The overall sample was over one million applicants. Limiting this report to only SAT total scores for GPA values between 2.5 and 4.5 for larger racial/ethnic groups reduced the sample to 628,946 (Table 1) with the smallest cell size having 149 cases.

Relationship Between Mean Test Scores and GPA

Figure 1 depicts the monotonic relationship between GPA values at the 10th of a point and SAT total scores for all groups and subgroups. Although not included for space purposes. Because the test point increments are small (1/10th of a point), and given the consistent nature of these relationships, if the test scores were unbiased, one should see a relatively random distribution of test means between and across groups for any given GPA score (e.g. 3.1). Each group should have an approximately equal chance of having higher or lower test scores than any other group at a given GPA point. However, the results show consistent differences favoring males and whites at almost all GPA scale points against every group, including Asians. This effect among the most affluent group, Asians, likely results from the cultural and second-language biases noted above. The white advantage occurs even for the SAT quantitative subtests against Asians until GPA reaches 4.0. Further, the mean differences between groups range from lows of about 20 points, to highs of about 140 points; averaging 75 for males over females (Figure 2). As was noted earlier, ETS (2001) suggests: "The user can be reasonably confident that a score difference of around 60 points or more indicates a true difference in ability between two test takers."

Figure 1
Relationship Between Test Scores and High School GPA by Race/Ethnicity
(N=628,948)

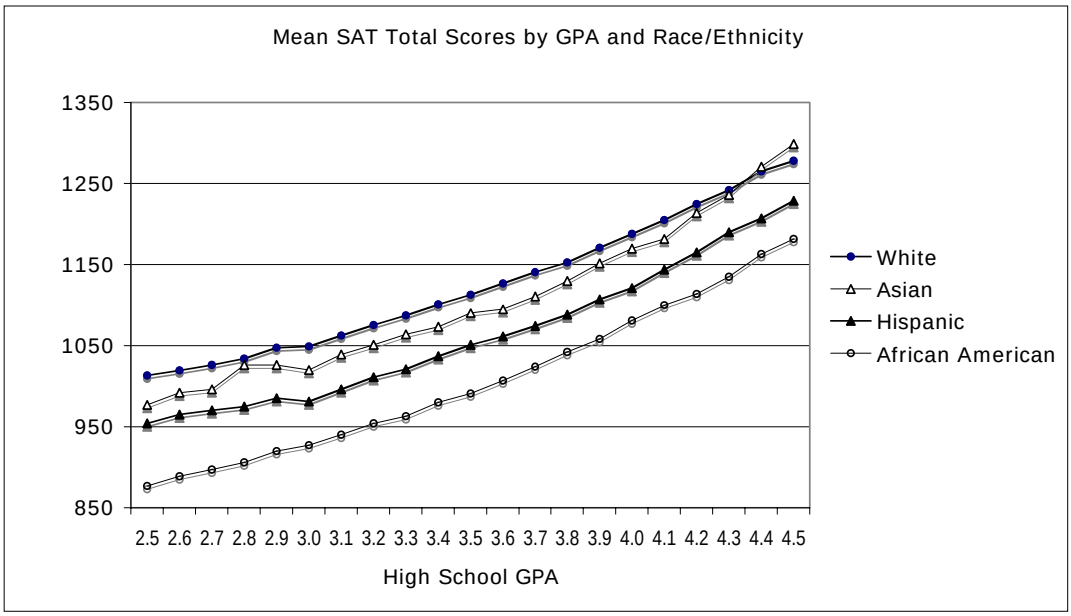
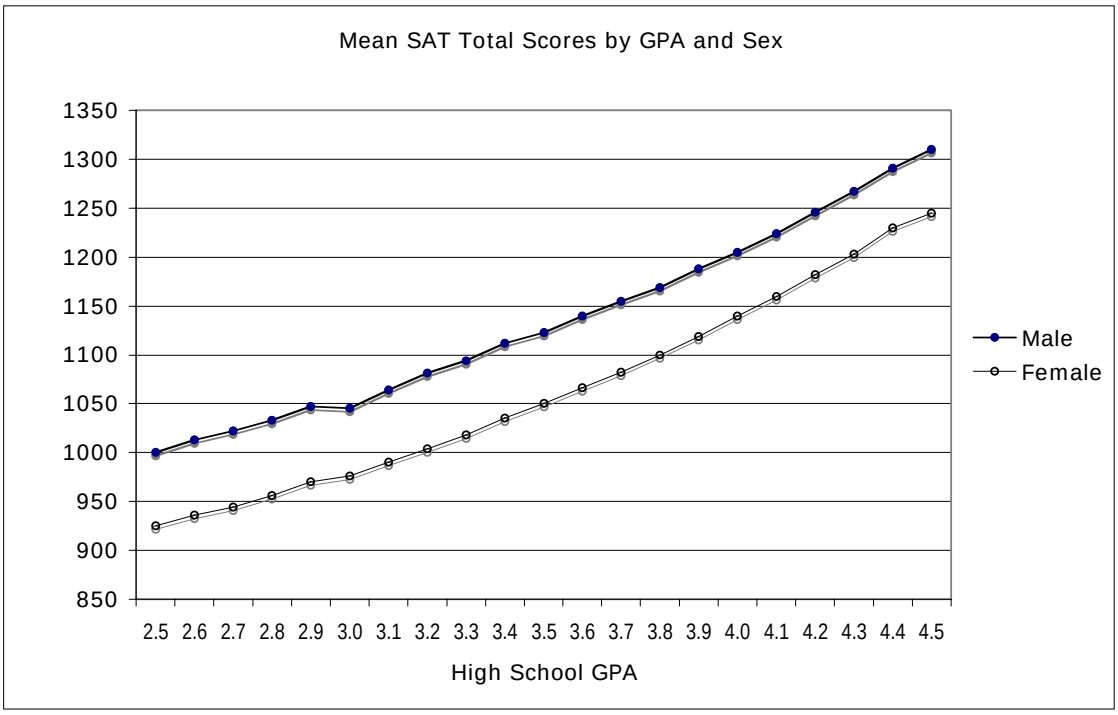


Figure 2 shows the consistent relationship between GPA and SAT scores for females versus males. As one can see the gap becomes smaller as GPA increases, just as it did across race/ethnicity in Figure 1.

Figure 2
Relationship Between Test Scores and High School GPA by Gender (N=628,948)



GAPS Become Smaller as GPA Values Increase

As student GPA values increase, the advantage favoring whites and males tends to become smaller. This may indicate increasing affluence as GPA increases, due perhaps to more time for schoolwork. At the 4.3 GPA level, Asian students (the most affluent) begin scoring higher than white students on the SAT. Interestingly, Asians do better on the quantitative than the verbal section, matching whites by the 4.0 GPA level. The same trends occur for both subtests with the match for verbal occurring at 4.5. Given the 2001 statement by ETS about how a difference of 60 points represents a "true" difference in ability, one must realize looking at Figure 1 and Figure 2 that if their statement is accurate, African American and Female students always have "less" ability than whites and males, and that only at score points of 4.3 to 4.5 do Hispanic students have no "true" disadvantage in ability.

Table 1 gives detailed values for all comparisons discussed in this paper.

Table 1
Mean Scores on Total SAT Tests for All FTIC Applicants, Summer 1997 through Spring 2006

HS	Asian				African American				Hispanic				White			
GPA	Female		Male		Female		Male		Female		Male		Female		Male	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
2.50	218	946	314	998	2,347	843	2,526	909	1,299	914	1,625	986	4,313	971	6,108	1042
2.60	232	942	367	1023	2,787	857	2,523	925	1,612	931	1,928	994	5,326	978	7,019	1050
2.70	325	951	472	1027	3,082	866	2,582	934	1,882	931	2,117	1004	6,125	986	7,682	1058
2.80	330	988	550	1049	3,524	876	2,773	945	2,302	939	2,363	1011	7,606	997	8,916	1066
2.90	467	987	631	1055	3,785	889	2,779	963	2,621	952	2,666	1017	8,784	1010	9,616	1081
3.00	666	989	748	1047	4,532	903	3,006	964	3,561	955	3,130	1011	10,131	1016	10,550	1082
3.10	667	999	847	1070	4,682	915	2,961	978	3,680	967	3,027	1032	10,994	1028	10,952	1097
3.20	723	1011	811	1087	4,446	930	2,540	996	3,640	980	2,951	1048	11,253	1041	10,528	1111
3.30	811	1028	951	1095	4,613	942	2,384	1005	3,872	994	2,939	1057	12,500	1053	11,512	1124
3.40	838	1042	868	1103	4,310	959	2,065	1022	3,839	1006	2,824	1079	12,641	1068	11,018	1140
3.50	906	1062	887	1119	4,010	967	1,907	1040	3,793	1023	2,737	1089	12,871	1083	10,803	1148
3.60	977	1063	950	1127	3,758	988	1,625	1053	3,814	1037	2,496	1098	13,001	1096	10,398	1165
3.70	1,017	1084	866	1142	3,172	1005	1,363	1067	3,496	1049	2,164	1114	12,247	1111	9,548	1180
3.80	1,072	1102	941	1163	2,903	1023	1,150	1091	3,406	1065	2,136	1126	12,830	1127	9,310	1190
3.90	1,042	1126	852	1185	2,625	1043	995	1098	3,249	1087	1,948	1140	12,364	1144	8,911	1210
4.00	1,125	1140	800	1211	2,080	1067	800	1118	2,719	1099	1,593	1158	11,251	1162	7,555	1226
4.10	1,013	1160	730	1211	1,875	1086	657	1140	2,377	1121	1,269	1187	10,609	1181	6,870	1241
4.20	908	1192	660	1244	1,333	1097	447	1165	1,906	1143	976	1209	8,713	1203	5,504	1259
4.30	929	1217	599	1266	1,116	1126	374	1161	1,578	1170	802	1230	8,483	1219	5,089	1281
4.40	779	1243	521	1314	763	1154	232	1193	1,199	1189	654	1241	7,099	1243	4,029	1302
4.50	583	1277	424	1330	539	1175	149	1206	879	1208	431	1273	4,951	1255	2,802	1319

Summary, Conclusion and Implications

Detailed analyses in this discussion was limited to SAT total scores, although similar consistent differences occurred for both SAT Subscores and ACT scores. The following conclusions appear consistent with the outcomes of these extensive data analyses:

- All groups and subgroups show a monotonic relationship between high school GPA values to the 10^{th} of a point and test score values.
- Given that one should expect a random distribution of differences at the same 10^{th} of a GPA level, these analyses indicate that consistent biases favoring whites and males occur in test scores.
- In most situations, the gaps favoring whites or males are above the 60 points, which ETS (2001) claims represents a "true" difference in ability. This "true" difference occurs where minority and female students are exhibiting exactly the same level of academic performance in school as measured by GPA.
- The gaps favoring whites and males tend to decrease as GPA values increase.

The commonly used excuse for research on tests, restriction of range, is not relevant for this study due to the large and representative nature of the sample (all SUS applicants, not merely admitted). Perhaps the most vital point regarding these analyses is that GPA is the criterion of success in school, whether high school or college. Standardized test scores, no matter how high, do not earn a degree, the certificate verifying academic success or failure. Thus, GPA must be considered the "true" score, with test scores being merely a proxy. Finally, it is important to realize that GPA tends to be a highly reliable estimate of academic performance and certainly, next to affluence, the strongest predictor of success in college (Florida SUS, 1995; Mortenson, 1999, 2000, Adelman, 2006). A student's high school or college GPA derives from multiple observers (teachers/professors), in multiple disciplines, over an extended period of time (four years in high school, four to six years in college). As a result, GPA must be considered a more valid estimate of academic performance than the point-in-time, strictly abstract estimates that standardized tests provide.

In summary, these results are quite consistent as is the presence of consistent biases favoring whites and males. This is important when viewed within the context of the vast empirical research literature showing how ineffective standardized tests are as predictors of college success beyond first semester grades. Increasing use of higher test scores as an admissions criterion appears to be a factor, if not the primary factor in the increasing class stratification of higher education. On the topic, Astin & Oseguera (2004, p. 335) note: "…students' socioeconomic status directly affects their chances of enrolling in a highly selective institution…"

As this study shows, the increasing use of test scores as an essential criterion for entry to prestigious institutions introduces an unjust and socially regressive bias favoring males, whites and the affluent (Heller & Marin, 2002). This reflects poorly on the integrity of higher education as institutions continue to perpetuate discriminatory practices that reduce access for minorities and females to the most prestigious academic conduit. The problem intensifies as one moves up the ladder of prestige and its effects are noted by Haycock & Gerald (2006, p. 7): "…the 50 flagship universities now look less and less like America—and more and more like 'gated communities of higher education.'" Further (p. 5): "…the highest achieving students from high-

income families-those who earned top grades, completed the full battery of college prep courses, and took AP courses as well-are nearly four times more likely than low-income students with exactly the same level of academic accomplishment to end up in a highly selective university." These recent findings extend the research reported by Mortenson, 1990; Heller & Marin, 2002, and Astin & Oseguera, 2004. Sacks (2007) suggests that one factor in the "gated community" effect is "…for those born with silver spoons, standards are relaxed and hands are held through every stage of the admissions process at selective colleges. In return for such favors, wealthy parents and donors lavish such institutions with money." The current study indicates that probably the most important factor behind this "gated community" effect is the use of test scores that are biased against all minorities, females and the less affluent.

Rankings such as U.S. News and World Reports 'America's Best Colleges," apparently increases the use of tests to augment an institution's perceived selectivity. Haycock & Gerald note (p. 3): "Rated less for what they accomplish with the students they let in than by how many students they keep out, many of these flagship institutions have become more and more enclaves for the most privileged of their state's young people." Astin & Oseguera (2004, p. 322) record the ultimate outcome: "…the most sought-after employers and most graduate and professional schools favor the graduates of prestigious institutions in their recruitment practices (Henson, 1980)." Further, because most faculty come from elite colleges (e.g. Flagship), this discrimination perpetuates the historic academic status quo of comparatively affluent, white and male faculty.

Perhaps even worse is the recent proliferation of high stakes tests in public K-12 education. All such tests exhibit precisely the same types of biases explicated herein and assure that academically qualified minorities and women are less likely to attain a high school diploma than are equally qualified whites and males. Further, the costs of even the most well designed such tests (e.g. Florida's FCAT) in terms of reduced instructional time and test preparation rather than curriculum-oriented instruction is simply vast, consuming perhaps 10 percent of the total K-12 funding (Micceri, 2001). Again, such effects are greatest for less affluent schools (lowest scoring on the tests), which again fosters advantages for the more affluent before they ever reach college.

A point about so-called "grade inflation." Statements regarding "grade inflation" usually occur when grades are higher, but test scores are not. Even Astin & Oseguera (2004, p. 336) state: "As grade averages become more and more inflated, they are given decreasing weight in the admissions process." It appears interesting to this observer that people would rather attribute higher grades to "inflation" than to what a rational individual would consider the more likely and powerful causes that, given the obvious financial benefits of higher education in a society where a college education has become almost a *sine qua non* for a living wage, both students and teachers are working harder and becoming more effective over time to increase students' chances of getting merit scholarships and admission to "better" higher education institutions (e.g. 4-year schools).

A Call for Change

Study after study has shown that neither for graduates nor undergraduates do standardized tests provide much useful information beyond that provided by GPA regarding a student's likelihood of success in college. For undergraduate students, the three factors that prove the best predictors of academic success are affluence, GPA and the rigor of high school

coursework a student successfully completes (Adelman, 2004, 2006). Hopefully, the current research provides adequate evidence regarding how biased, unethical and socially regressive the use of standardized test scores as admissions criteria truly are and will result in a move away from the use of such repressive and costly criteria.

References

Adelman, C. (2006). *The toolbox revisited: Paths to Degree Completion From High School Through College*. Washington, D.C.: U.S. Department of Education, 2006.

Adelman, C. (2004). *The empirical curriculum: Changes in Postsecondary Course-Taking, 1972-2000*. Washington, DC: U.S. Department of Education.

Astin, A.W. & Oseguera, L. (2004). The declining "equity" of American higher education. *The Review of Higher Education*, 27:3, pp. 321–341.

Astin, A.W. (1993). *How good is your institution's retention rate?* Unpublished manuscript, University of California, Los Angeles.

Bangura, A.K. (1995). *Hold everything!: Emerging Problems in Institutional Accountability for Retention and Graduation*. (ERIC Document Reproduction Service No. ED 410845).

Bieker, Richard F. (1996). Factors Affecting Academic Achievement in Graduate Management Education. *Journal of Education for Business*; v72 n1 p42-46 Sep-Oct 1996.

Borman, K., Workman, C., Miller, J. & Micceri, T. (2006). *The Role of the Community College in Contemporary Society: The Florida context*. Paper presented at the Annual American Anthropological Association Conference, San Jose, CA, November 15-20.

Bracey, G.W. (1997). *Setting the Record Straight: Responses to Misconceptions about Public Education in the United State*s. Association for Supervision and Curriculum Development. Alexandria, Va. Retrieved August, 1999 from http://www.ascd.org/readingroom/books/bracey97book.html, or http://tinyurl.com/3b4y7e.

Breland, H.M. & Minsky, S. (1978). *Population validity and college entrance measures*. Princeton, NJ: Educational Testing Service, 1978.

CDC (2006). U.S. Census Bureau Census Download Center. Retrieved August 28, 2006 from http://tinyurl.com/guad6, or http://factfinder.census.gov/servlet/DownloadDatasetServlet?_lang=en.

Chernyshenko, O S. & Ones, D.S. (1999). How selective are psychology graduate programs? The effect of the selection ratio on GRE score validity. *Educational and Psychological Measurement*; v59 n6 p951-61 Dec 1999.

Commonfund (2006). *2006 HEPI – Higher education price index*. Commonfund Institute. Retrieved September 1, 2006 from http://www.commonfund.org/Commonfund/Investor+Services/HEPI.htm?m=h&AA=2, or http://tinyurl.com/jqhch.

Crouse, J. & Trusheim, D. (1988). *The Case Against the SAT*. Chicago: University of Chicago Press.

Elert G. (1992). The SAT, aptitude or demographics? *E-World*, Retrieved December, 1999 from. http://hypertextbook.com/eworld/sat.shtml

ETS (2001). *Test Characteristics of the SAT I: Reliability, difficulty levels, completion rates*. Retrieved February, 2002 from http://www.collegeboard.org/sat/cbsenior/stats/stat002.html or http://tinyurl.com/38xu4r.

FairTest (2006). *Gender bias in college admissions tests*. The National Center for Fair and Open Testing. Cambridge, MA. Retrieved June, 2006 from http://fairtest.org/facts/genderbias.htm..

Florida SUS (1995). *A study of admissions criteria*. Technical Report, Florida Board of Regents, Tallahassee, FL.

French Jr., John R. P. (1967). "Quantification of organizational stress." *Managing Organizational Stress: Proceedings of the Executive Study Conference*: November 29-30, 1967. Princeton, NJ: Educational Testing Service, 1968: 5-41.

Gibson, R. (2001). *The fascist origins of the SAT test*. BBC-NEWS, April, 2001. Retrieved April 8, 2001 from http://www-rohan.sdsu.edu/~rgibson/SATFascistOrigins.htm or http://tinyurl.com/3ajnhx

Goldin, C. & Katz, L.F. (1999). The shaping of higher education: The formative years in the United States, 1890 to 1940. *The Journal of Economic Perspectives*, 13:1, pp. 37-62.

Haycock, K. & Gerald, D. (2006). *Engines of inequality: Diminishing equity in the nation's premier public universities*. Technical Report of the Education Trust, Retrieved November 20, 2006 from http://www2.edtrust.org/NR/rdonlyres/F755E80E-9431-45AF-B28E-653C612D503D/0/EnginesofInequality.pdf.

Heller, D.E. & Marin, P. (eds.) (2002). *Who should we help? The negative social consequences of merit aid scholarships*. The Civil Rights Project, Harvard University. Retrieved Febuary 15, 2007 from http://www.civilrightsproject.harvard.edu/research/meritaid/fullreport.php

Hodgkinson, H. (1999) *CONNECTION: New England's journal of higher education and economic development*. September 22, 1999 - Retrieved October, 2000 from http://www.nebhe.org/bkissues/Sum99.html

House, J. D. (1998). *Gender differences in prediction of graduate course performance from admissions test scores: An empirical example of statistical methods for investigating prediction bias*. AIR 1998 Annual Forum Paper. (ERIC Document Reproduction Service No. ED 424810).

House, J. D.; Gupta, S.; Xiao, B. (1997). *Gender differences in prediction of grade performance from graduate record examination scores*. (ERIC Document Reproduction Service No. ED 415280).

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. Farrar, Straus and Giroux, NY.

Lemann, N. (1999a). Behind the SAT. *Newsweek*, 9/6/1999, pp. 52-56.

Lindle, Jane Clark; Rinehart, James S. (1998). *Emerging issues with the predictive applications of the GRE in educational administration programs: One doctoral program's experience*. (ERIC Document Reproduction Service No. ED 424632).

Micceri, T. (2001). *Why do we waste so many resources on incredibly costly and practically worthless standardized tests?*. Internal Technical Report, Office of Budget and Policy Analysis, USF, Retrieved September 9, 2005 from http://isis.fastmail.usf.edu/surveys/Studies/.

Micceri. T. (2005). *An attempt to identify comparatively supportive and non-supportive environments for underrepresented minorities and females in SUS colleges of engineering*. Internal Technical Report. AAREA. University of South Florida, Tampa, FL., December, 2005. ERIC Document: ED497450.

Morrison, T. & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the graduate record examination with graduate grade point average representing the criterion of graduate success. *Educational and Psychological Measurement*; v55 n2 p309-16 Apr 1995

Mortenson, T.G. (2000). *Opportunity for higher education in Florida in the human capital economy*. Presentation to the Florida Postsecondary Education Planning Commission.

Mortenson, T.G. (1999). The impact of financial aid on college retention for minority students. *Postsecondary Education Opportunity*, Oskaloosa, IA, 52577.

Mortenson, T.G. (1990). The reallocation of financial aid from poor to middle income and affluent students, 1978 to 1990. *ACT Student Financial Aid Research Report Series 90-2*.

MSCHE (1997). *Relationship between high school and college performance by Maryland students. Student outcome and achievement report*. Maryland State Commission on Higher Education, 16 Francis Avenue, Annapolis, MD 21401; phone: 410-974.2971.

Murphy, M. (2000). *Predicting Graduation: Are Test Score and High School Performance Adequate*. Paper presented at the AIR Annual Forum, Cincinnati, OH, May 21-24, 2000.

Onasch, C. (1994) – *Undergraduate grade point average and graduate record exam scores as predictors of kength of enrollment in completing a master of science degree*. (ERIC Document Reproduction Service No. ED 375739).

Palmer, J.D. (2004). *GRAPEVINE Survey of state higher education tax appropriations for fiscal year 2004*. Center for the Study of Education Policy, Illinois State University, Normal, IL.

Paszczyk, S. L. (1994). *A comparative analysis of ACT scores and final GPAs of Chicago state university undergraduate students*. (ERIC Document Reproduction Service No. ED 370519).

Raven, J.C. (1939). The R.E.C.I. series of perceptual tests: An experimental survey. *British Journal of Medical Psychology*. 18, pp. 16-34.

Robbins, J. (1987). *Diet for a new America*. Stillpoint Publishing, Walpole, NH.

Rosser, P. (1989). *The SAT gender gap: Identifying the causes*. Center for Women's Policy Studies, Washington, D.C.

Sacks, P. (2007). How colleges perpetuate inequality. *The Chronicle of Higher Education*. 53:19, p. B9.

Sahadi, J. (2006). *Wealth gap widens: Chasm between wealthiest households and everyone else has grown more than 50 percent since the early 1960s*. CNNMoney.com. August 29 2006. Retrieved December 28, 2006 from http://money.cnn.com/2006/08/29/news/economy/wealth_gap/index.htm?postversion=2006082913 or: http://tinyurl.com/37gjed

SHEEO (1997). *State higher education appropriations*. State Higher Education Executive Officers. Boulder, CO. Retrieved September 1, 2006 from http://www.grapevine.ilstu.edu/Appropriations1996-97.pdf, or: http://tinyurl.com/3ag254.

Sternberg, G. & Williams, J. (1997). Does the graduate record examination predict meaningful success in the graduate training of psychologists? A case study. *The American Psychologist*, 52, 630.

Stricker, Lawrence J.; Rock, Donald A. (1993). *Examinee background characteristics and GRE general test performance*. GRE Board Research Report No. 89-07R. (ERIC Document Reproduction Service No. ED 385589).

Tisue, Linda J.; Whitaker, Dianne P. (1999). The GRE: A barrier to women returning to graduate school? *Journal of Educational Opportunity*; v17 n2 p31-43.

Underbakke, M. (2000). Personal Communication.

Vogel, M. (2006). The college admissions squeeze: Florida colleges. *Florida Trend*. August 31, 2006.

Waugh, G., Micceri, T. & Takalkar, P. (1994). *Using ethnicity, SAT/ACT scores, and high school GPA to predict retention and graduation rates*. Paper Presented at the Annual FAIR Conference, Orlando, FL, June 15-17, 1994.

Webster, B.H, Jr., & Bishaw, A. (2006). *Income, earnings, and poverty data from the 2005 American community survey*. U.S. Census Bureau, American Community Survey Reports, ACS-02, U.S. Government Printing Office, Washington, DC.

WICHE (2003). *Knocking at the college door: Projections of high school graduates by state, income, and race/ethnicity*. Florida report, Retrieved January 15, 2007 from http://www.wiche.edu/policy/Knocking/1988-2018/profiles/fl.pdf or: http://tinyurl.com/36gp6g.

Witte, G. & Henderson, N. (2004). Wealth gap widens for Blacks, Hispanics: Significant ground lost after recession. *Washington Post*, October 18, 2004; Page A11. Retrieved December 28, 2006 from http://www.washingtonpost.com/wp-dyn/articles/A40455-2004Oct17.html or: http://tinyurl.com/yaxokd.

Xiao, B. (1998). *Factors influencing master's degree attainment in business, engineering, health and human sciences and visual and performing arts*. Paper Presented at the AIR Annual Forum, Minneapolis, MN, May 17-20, 1998.