Schooland: Drivers of Standardized Test Performance and Graduation Rates[1]

1. Introduction and Key Findings

Determining what drives pupils' success in school is a key concern for policymakers and parents alike. Success in school is a positive predictor of a person's future income, health, and commitment to one's family and community. However, drives of school success can vary significantly across the nation.

Listed in this report are the key determinants of ACT scores and graduation rates for students at Schooland High. The main findings are:

- As expected, **overall academic performance during high school is good predictor of ACT scores and graduation rates. Increasing one's Grade Point Average (GPA) by 10% is expected to increase one's ACT score by 1%.** This value may be slightly understated, as students with a lower GPA have a smaller probability of taking the ACT altogether. **The relationship between GPA and graduation rates is positive, though smaller. Each point increase in the student's GPA (say, increasing from 2.5 to 3.5) increases that student's probability of graduating by 4.7%**.
- **Gender plays a statistically significant role in determining a student's ACT score**, even after controlling for other characteristics. **Being a female increases one's expected score by 7.9%, but has no significant impact on probability of graduating**.
- Unlike most counties in the nation, **household income has no statistically significant impact on either graduation rates or standardized test performance.** For students attending Schooland High, household income is completely uncorrelated with school performance – if anything, it is slightly negatively correlated.
- **Ethnicity is a poor predictor** of ACT scores and graduation rates for Schooland students. Taken as two groups, there is **no expected score difference between white and non-white pupils**. However**, Asian students tend to have lower graduation rates and ACT scores –** 6.8% and 1.9% lower as compared to whites, respectively.
- **Participation in specific school programs are good predictors** of both ACT scores and graduation rates**.** Taking an **Advanced Placement class increases one's expected ACT score by 10.2% while raising one's probability of graduating by 8.9%**. In contrast, **being an athlete decreases a student's predicted ACT by 8.8% while cutting her probability of graduating by 6%**. It should be noted that there could be selection bias in those estimates. It is not necessarily the case that AP classes or Athletics Programs increase or decrease ACT scores or graduation rates per se. It is possible that students who would have higher ACT scores regardless of whether they enrolled in AP classes choose to take them in the first place. This selection bias could be overstating the effects of such programs[2].

2. Data Adjustments and Description

The prepared dataset merges school performance, program participation, and demographic information with ACT scores results. All information is uniquely identified by a numerical student code, protecting pupils' privacy.

Minor adjustments in the data were made. They include:

- The calculation of a Grade Point Average following standard methods, ranging from 0 through 4[3].
- For pupils who took more than one ACT, only the latest result was taken into account. We stored the number of times each student took the standardized test.
- For simplicity, we changed the scale of household income to thousands of dollars per household and calculated income quartiles.
- Twenty students had their ages wrongly stored. To keep data consistency, we randomly assigned new ages for them, using the sample's age mean and standard deviation.
- We did not use the performance index listed in the dataset, since the data were available for only a fraction of total Schooland High population.
- We created log-transformed variables of ACT scores, GPA, and household income, as well as dummies for the binary variables (see Appendix for further description).

[1] Prepared by Carlos Góes.

[2] In a future statistical study, a solution to this selection bias would be applying the Heckman Correction.

[3] The numerical equivalents of letter grades are as follows: A = 4, B = 3, C = 2, D = 1, F = 0. After performing a numerical conversion, the GPA was calculated as the arithmetic average of each student's grades.

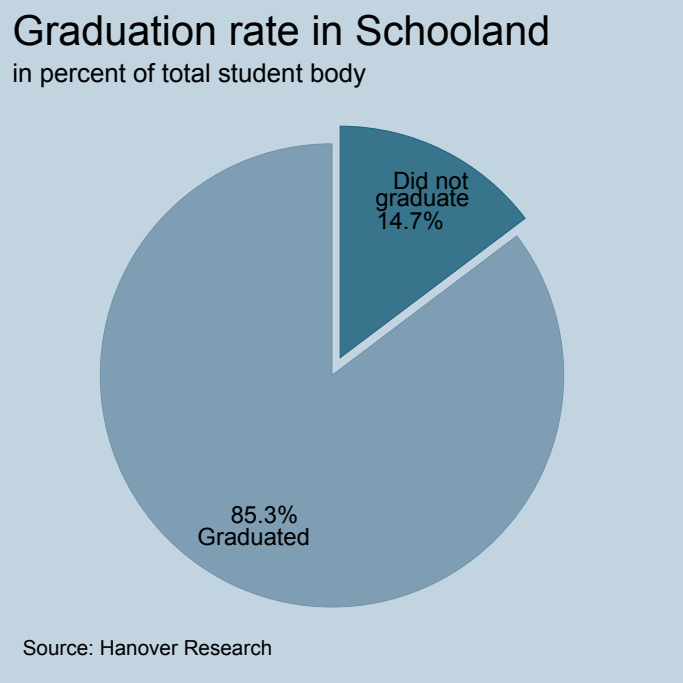
Table 1 below summarizes the demographics and school program enrollment participation rates.

- **Schooland High is ethnically diverse**, with minority students outnumbering white students.
- **There is some gender unbalance, with males outnumbering females**. However, there is some missing information with regard to gender, with more than 18% of students failing to report their gender.

Table 1: School Demographics, Performance, and Program Enrollment

Education Programs		Rate
Tried the ACT		80%
Participation in Athletics Program		35.9%
Participation in Advanced Program		41.7%
Participation in Special Education Program		9.5%
Gender		
Female		38.1%
Male		43.2%
Undeclared		18.7%
Ethnicity		
African American		28.6%
Asian		19.2%
White		41.9%
Other		10.3%
School-Related Information	**Average**	**Std Deviation**
Grade Point Average (0 - 4)	2.4	0.5
Performance Index (0 - 1)	0.7	0.2
ACT Test Score (1 - 36)	23.5	2.8
# of ACTs taken	1.6	0.8
Other demographics	**Median**	**Std Deviation**
Age	0.0	0.6
Household income, $000	50.9	18.2

- **Average GPA and ACT scores in Schooland High fall slightly above the expected averages** of 2.0 and 20, respectively.
- **Median household income is 50.9k dollars,** slightly above the national average.



3. Method

We use **two standard methods** to determine the drivers of ACT scores and graduation rates in Schooland, namely **multivariate linear regression analysis** and **logistic regression analysis**.

A **linear regression** (or "ordinary least squares") simply expresses the expected (or "average") increase in a dependent variable (in our case, ACT scores) due to changes in predictors (in our case, GPA, gender, ethnicity, etc.)[4]. The way we interpret these estimates is by posing questions such as: "*How many more points is John Doe expected to achieve, holding all other factors constant, if he takes at least one Advanced Placement course?*"

A **logistic regression** estimates the changes in probability of a certain value being equal to 100%. We first categorize those students who have graduated and those who have not. We then estimate the additional (or "marginal") increase in their probability of graduation based on certain individual characteristics (e.g. GPA)[5]. The way we interpret these estimates is by posing questions such as: "*How much more likely is median student John Doe to graduate if he is an athlete (versus being a non-athlete)?*" or "*How much an extra $1,000 of household income increases the median student John Doe likelihood to graduate?*"

[4]More formally, the model we used was a standard cross-section model $\hat{y} = \hat{c} + X'\hat{\phi} + \hat{e}$, where c is a constant, X' is the transposed vector of regressors, ϕ is a vector or coefficients, and $\hat{e}$ are the residuals.

[5]More formally, the model we used was a standard logistic regression $\Pr(y = 1|X') = 1 / [1 + \exp(1/(\hat{c} + X'\hat{\phi}))]$.

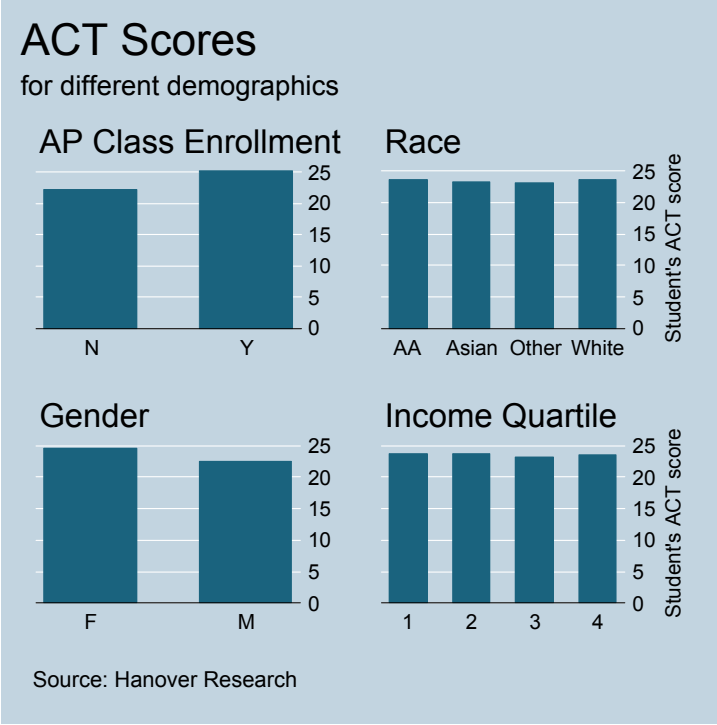
4. Results

4.1. ACT Scores

Before estimating regressions, we eyeballed possible simple relationships between general student demographics and the variables that we are trying to explain. Such visualization can give us some intuition into which characteristics are the most associated with changes in ACT scores and help us to judge our final results.

Looking at the figure below, one can see that **ACT scores seem largely disassociated with ethnicity and income.** In fact, the mean values of ACT scores for whites and non-whites are not statistically different[6].

The same is not true for other characteristics. **Women tend to score higher in standardized tests**. **So too do students who have enrolled in Advanced Placement courses.** By contrast, those **students who take part in Athletics programs tend to have lower scores**. The picture below summarizes ACT scores broken down by some demographic characteristics.



The expected positive correlation between GPA and standardized test score does hold. But Schooland High seems to be a national outlier in the sense that **household income is largely uncorrelated both with GPA and ACT scores**.

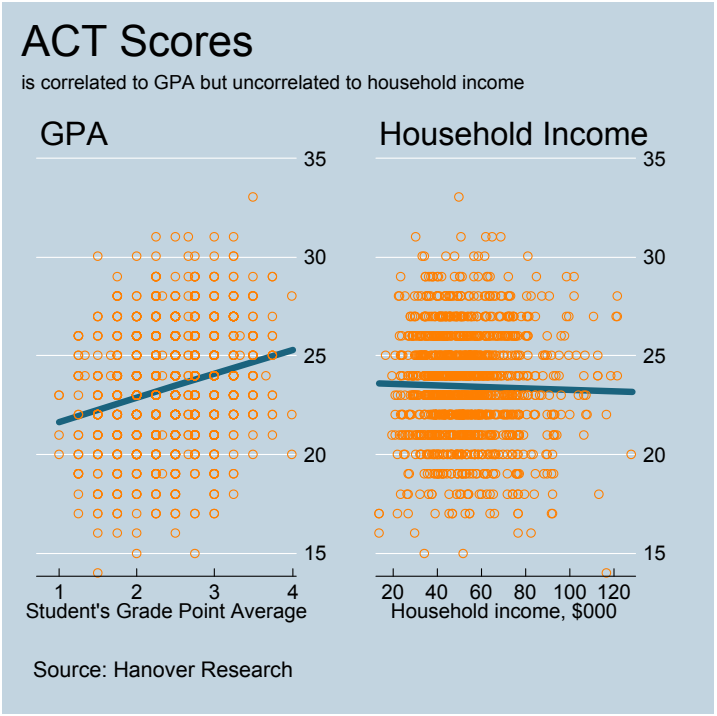


Table 2 below shows the results of the multivariate linear regressions. Columns (1) and (3) illustrate the same relationships, but in a different fashion. **Column (1)** shows expected percentage changes in ACT scores while **column (3)** shows expected point changes in ACT scores.

In columns (2) and (4), we dropped from our sample the students who did not declare their gender, to see if the results hold. **The results are broadly robust to this sample change**.

[6] Mean comparison tests between whites and non-whites scores fail to reject the null hypothesis that the difference between the scores is different than zero.

Table 2: Predictors of ACT Score

In columns (1) and (2), the coefficients relate to predicted percent changes in ACT score while in column (3) and (4) they predict changes in the actual score. In (2) and (4), we dropped students with undeclared gender, to see if the results from baseline result hold.

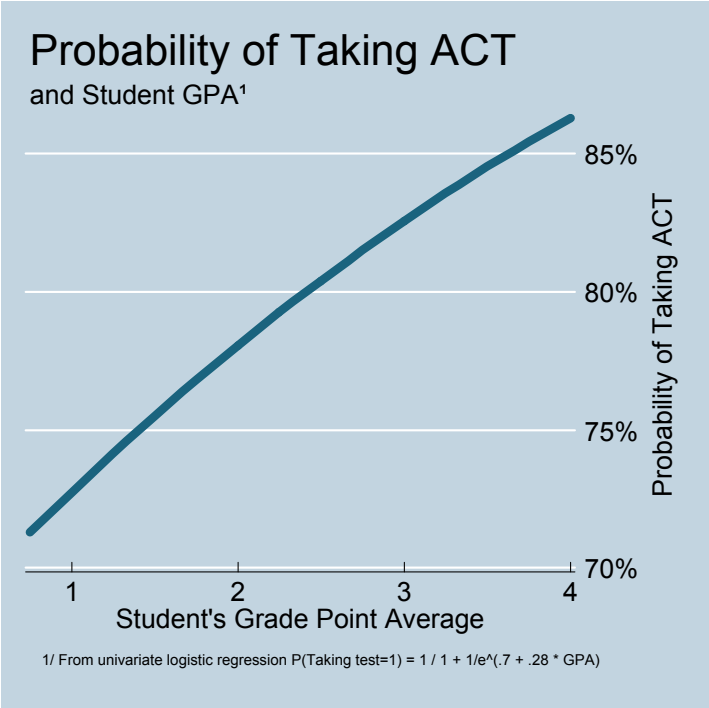
	(1) Dependent Variable: Natural Log ACT Score	(2) Dependent Variable: Natural Log ACT Score	(3) Dependent Variable: Actual ACT Score	(4) Dependent Variable: Actual ACT Score
Participation in Advanced Program	**0.123*** (0.006)**	**0.126*** (0.006)**	**2.889*** (0.134)**	**2.958*** (0.140)**
Graduated	0.014 (0.009)	0.015 (0.009)	0.313 (0.194)	0.317 (0.203)
Participation in Special Education Program	0.009 (0.010)	0.007 (0.011)	0.198 (0.225)	0.175 (0.233)
Participation in Athletics Program	**-0.088*** (0.006)**	**-0.084*** (0.007)**	**-1.992*** (0.142)**	**-1.897*** (0.148)**
Female Students	**0.079*** (0.006)**	**0.095*** (0.006)**	**1.825*** (0.134)**	**2.199*** (0.138)**
African American Students	-0.005 (0.007)	0.003 (0.007)	-0.128 (0.156)	0.026 (0.163)
Asian Students	**-0.019** (0.008)**	**-0.017** (0.008)**	**-0.440** (0.187)**	**-0.412** (0.186)**
Other Minority Students	-0.006 (0.011)	0.001 (0.012)	-0.120 (0.237)	0.046 (0.273)
Attempts Student Tried ACT	-0.003 (0.004)	**-0.008** (0.004)**	-0.050 (0.087)	**-0.173** (0.087)**
Student's Grade Point Average			**1.052*** (0.124)**	**1.005*** (0.129)**
Natural log of Student's Grade Point Average	**0.102*** (0.012)**	**0.099*** (0.013)**		
Household income, $000			-0.002 (0.004)	0.002 (0.004)
Natural log of Household income, $000	-0.006 (0.008)	0.005 (0.009)		
Constant	3.033*** (0.038)	2.977*** (0.040)	19.802*** (0.450)	19.360*** (0.489)
Observations	800	638	800	638
R-squared	0.548	0.606	0.553	0.611

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The most important statistically significant predictors of ACT score are GPA, participation in an Advanced Placement program, participation in an Athletics program, gender, being Asian, and (in the smaller sample) the number of times the student has taken the ACT.

- **GPA:** a 10% increase in GPA is expected to lead to a 1% increase in ACT scores.
- **AP**: a student who has taken AP classes is expected to have a 12.3% higher ACT score.
- **Athletics:** after controlling for other demographic variables, athletes tend to have 8.8% lower ACT scores.
- **Gender:** being female increases one's expected ACT score by 7.9%.
- **Ethnicity:** ethnicity is largely unrelated to ACT scores, **except for being Asian**, which decreases the expected score by 1.9%.
- Taking part in special education programs, ethnicity (other than Asian), and household income have no statistically significant impact on ACT scores.

It should be noted that **there could be selection bias in some of these estimates, underestimating the true impact that GPA has on ACT scores**. This happens because students with lower GPAs are up to 15% less likely to take the ACT in the first place, as seen in the chart below. If sitting for ACTs were mandatory, the estimated relationship between GPA and ACT scores could be higher.

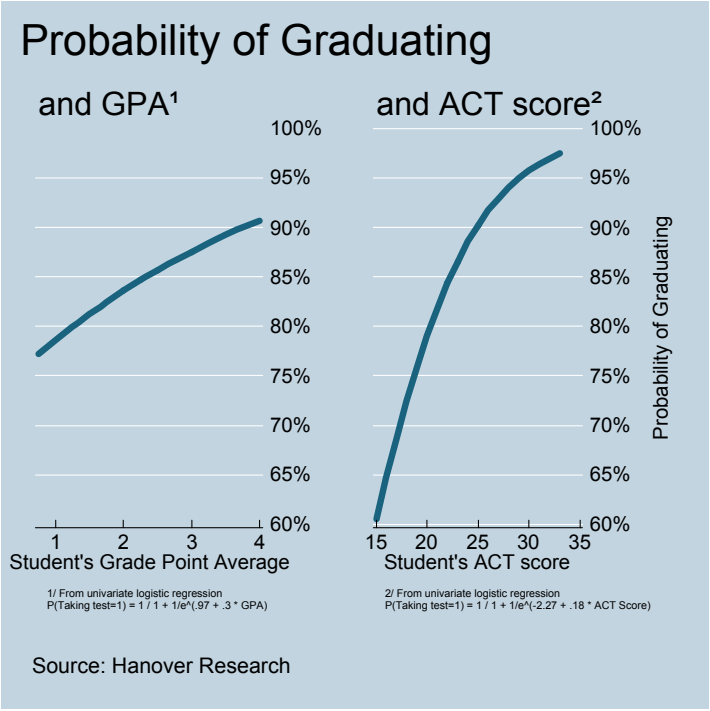


Similarly, **it is not necessarily the case that AP classes or Athletics Programs increase or decrease ACT scores or graduating rates per se**. It is possible that students who would have higher ACT scores regardless of whether they have enrolled in AP classes choose to take them in the first place. This selection bias could be overstating the effects of such programs[7].

[7] In a future statistical study, a solution to this selection bias would be applying the Heckman Correction.

4.2. Probability of Graduating

Insofar as GPA and ACT scores reflect ability, **individual ability seems to increase one's probability of graduating from high school**. But this relationship is non-linear, as shown in the graph below. Students with significantly below average ACT scores are much less likely to graduate. Once they reach the average ACT score (~23), any additional increase in ACT points will have a much smaller impact on their probability of graduating from high school.



From those univariate analyses, we can start to add layers of control, to see if the role of GPA and ACT scores (and, by extension, individual ability) still persists after controlling for other individual characteristics.

Table 3 below shows the results of the logistic regressions. Columns (1) and (2) show the univariate models used to plot the chart just above.

Column (3) includes all individual characteristics and both GPA and ACT, while Column (4), which we chose as our baseline model, drops ACT and keeps GPA.

In column (3), even though the individual coefficients for GPA, ACT score and AP participation are not statistically significant, they are jointly statistically significant (i.e, their combined value is statistically different from zero). Since we assume that individual characteristics drive both ACT and probability of graduating, we chose to treat the former as exogenous, and use model (4) as our baseline[8].

Table 3: Predictors of Probability of Graduating

Marginal effect of changing one unit of the independent variable (one extra point of GPA or $1,000 more in household income) on the probability of graduating, at their means. For binary variables (gender, race, and participation in AP, Athletics or Special ED programs) it means the increase in the probability of graduating due to, for instance, being African American or taking part in one of the programs.

	(1)	(2)	(3)	(4)
	Marginal Effects: Probability of Graduating			
Student's Grade Point Average	**0.041** **(0.020)**		0.033 (0.024)	**0.047**** **(0.024)**
Student's ACT score		**0.020***** **(0.004)**	0.010 (0.006)	
Household income, $000			-0.000 (0.001)	-0.000 (0.001)
Participation in Advanced Program			0.054 (0.036)	**0.089***** **(0.034)**
Participation in Special Education Program			0.029 (0.047)	0.034 (0.051)
Participation in Athletics Program			**-0.036** **(0.026)**	**-0.060***** **(0.023)**
Female Students			0.013 (0.029)	0.034 (0.030)
African American Students			0.002 (0.030)	0.001 (0.032)
Asian Students			**-0.058**** **(0.027)**	**-0.068**** **(0.029)**
Other Minority Students			-0.015 (0.040)	-0.018 (0.043)
Attempts Student Tried ACT			-0.003 (0.015)	-0.004 (0.016)
Observations	1,000	800	800	800

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The statistically significant predictors of one's probability of graduating are broadly similar to the predictors of one's ACT score. This is expected, since, as mentioned above, one's ACT score is itself a good predictor of his probability of graduating. **For the average student, the most important predictors are: GPA, participation in an AP program, being an athlete, and being Asian**.

[8] In our baseline marginal effects model (4), GPA, number of attempts and household income are set to their means, while the dummies for participation in an AP program, an Athletics program, and a Special Ed program are set to zero. We also set to zero the dummies regarding ethnicity and gender. This means that our "median individual" to whom the marginal effects apply is non-minority, non-female and does not take part in any academic program.

- **GPA:** for each additional GPA point earned, there is a 4.7% increased probability of graduating.
- **AP**: students who take AP classes are 8.9% more likely to graduate.
- **Athletics:** athletes are 6% less likely to graduate.
- **Ethnicity:** ethnicity is largely unrelated to one's probability of graduating, **except for being Asian**, which decreases one's likelihood of graduating by 6.8%.
- Taking part in special education programs, ethnicity (other than Asian), gender, household income, and the number of times one has taken the ACT have no statistically significant impact on one's probability of graduating.

Even after controlling for many individual characteristics, GPA remains a good predictor of probability of graduating. This is particularly important in Schooland High, since other demographics as ethnicity and income are not good predictors of standardized test performance or probability of graduating.

Parents, staff and policymakers who oversee Schooland High have all the more an incentive to work towards boosting GPA, since increasing school performance there will most likely lead to more meritocratic results than elsewhere.

5. Appendix – Summary of Variables

Table 4: Summary of Variables
This table summarizes and describes the variables contained in models used in the report

	Description	Type	# of Observations
Grade Point Average (GPA), [0 - 4]	Calculated form Letter Grades. The number equivalence of letter grades are A = 4, B = 3, C = 2, D = 1, F = 0. After converting to numbers, the GPA was calculated as the arithmetic average of each student's grades.	Continuous	1000
Household income, $000	Student's household income, in thousand of dollars	Continuous	1000
Performance Index, [0 - 1]	School calculated variable representative of student performance	Continuous	240
Graduated	Student graduated (1) or did not graduate (0)	Categorical	1000
Age	Student's age	Continuous	1000
Participation in Athletics Program	Student partipated in Athletics program (1) or did not partipate (0)	Categorical	1000

(continues in the next page)

Table 4: Summary of Variables (continued)

Participation in Advanced Placement Program	Student partipated in Advanced Placement program (1) or did not partipate (0)	Categorical	1000
Participation in Special Education Program	Student partipated in Special Education program (1) or did not partipate (0)	Categorical	1000
Female Students	Student is Female (1) or non-Female (0)	Categorical	1000
African American Students	Student is African American (1) or non-African American (0)	Categorical	1000
Asian Students	Student is Asian (1) or non-Asian (0)	Categorical	1000
White Students	Student is white (1) or non-white (0)	Categorical	1000
Other Minority Students	Student is belongs to other minorities (1) or not (0)	Categorical	1000
ACT Composite Test Score, [1 - 36]	Student's ACT test score	Continuous	800
Attempts Student Tried ACT	Number of times the student has taken the ACT	Continuous	800