Key Indicators of College Success: Predicting College Enrollment, Persistence, and Graduation

Kelly E. Godfrey

Haifa Matos-Elefonte

College Board

Background & Literature

- Plethora of research examining factors contributing to students' decision to enroll in college, their persistence, and graduation
- This research models the work of Johnson (2008)
  - Taking both student- and school-level characteristics
- And follows the approach of Adelman (2006)
  - focuses on the student's pathway through high school to enrollment, persistence, and ultimately graduation with a Bachelor's degree.

Data

- Student high school performance and higher education enrollment data were acquired from the Department of Education (DOE) from a large, diverse state located in the continental United States
- National assessment data were acquired from College Board's archives

Methods:
Hierarchical Generalized Linear Models (HGLMs)

- Hierarchical linear modeling is appropriate here because students are nested in high schools
- Two-level HGLMs predicting the three outcomes (enrollment, persistence, graduation) using a Bernoulli distribution
- HGLMs estimated through restricted penalized quasi-likelihood (PQL) estimation method
- Tests if student *and* school level variables can predict the outcomes

Methods: Variables

Student level variables

- Gender
- Limited English Proficiency status
- AP course taker flag
- Race (dummy coded into four variables)
- Free/reduced lunch status
- High school GPA
- SAT taker flag (enrollment only) or SAT Verbal & Math scores (persistence and graduation)
- 10th grade state assessment scores in Reading and Mathematics
- Percent of high school coursework at an honors, pre-International Baccalaureate (IB), IB, or AP level

School level variables

- Percent absent 21 days or more
- Percent free/reduced lunch status
- Mean ACT composite score
- Percent who took the ACT
- Mean SAT (Verbal + Math) score
- Percent who took the SAT
- Percent who took the PSAT
- Percent of teachers with a Master's degree or higher
- Teachers' average years of experience
- Total number of discipline referrals in 2001-2002 academic year
- Percent of class that are gifted
- Percent of class identified as ELL
- Number of AP courses taught at the school
- Student-teacher ratio
- Number of Full-Time teachers

Sample

Table 1: Basic Demographic Information

	Percentage of Sample
Gender	
Male	46.5%
Female	52.8%
Race/Ethnicity	
Asian	2.4%
Black	19.1%
Hispanic	15.7%
White	58.3%
Other	4.5%
Limited English Proficient	13.1%
Free/Reduced Lunch	37.5%
AP course takers	27.3%

Sample

Table 2: Percentage of Sample Enrolled, Persisted, and Graduated from a Public In-state Institution

Outcome	Percentage of Entire Sample
Enrolled in public, in-state institution Fall 2002	41.3%
Persistence to Fall 2003	30.2%
Graduation with Bachelor's within 5 years	12.1%

Sample

Table 3: Percentage of Free/Reduced Lunch and Limited English Proficient students who Enroll, Persist, Graduate from a Public In-state Institution

Outcome	Percentage of F/R Lunch	Percentage of LEP
Enrolled in public, in-state institution Fall 2002	33.0%	41.6%
Persistence to Fall 2003	21.9%	30.6%
Graduation with Bachelor's within 5 years	5.4%	7.3%

Table 4: Percentage of SAT Taking and AP Participating Students who Enroll, Persist, Graduate from a Public In-state Institution

Outcome	Percentage of SAT Takers	Percentage of AP Takers
Enrolled in public, in-state institution Fall 2002	55.9%	55.8%
Persistence to Fall 2003	44.8%	47.8%
Graduation with Bachelor's within 5 years	21.4%	28.8%

Results: Enrollment

Enrollment =1 if a student enrolled in a 2 or 4 yr college in the fall semester immediately following graduation from high school (Fall 2002)

Table 5: Empty Logistic Multilevel Model Predicting Enrollment

Fixed Effects	Coefficient (SE)	Odds Ratio	t (df)	*p*
Intercept (γ_{00})	-0.61 (0.04)	0.55	-14.75 (382)	0.00

Random Effects	Variance	df	Chi square
τ_{00}	0.56	382	8487.87 (0.00)

The estimated probability of enrollment for students in this sample is .35 (calculated as exp(-.61)/(1 + exp(-.61) = 0.54/(1 + 0.54) = 0.35)

The intraclass correlation coefficient for this model is .15 (calculated as ICC = τ_{00}/(τ_{00}+3.29) = .56/(.56+3.29) = .15)

Results: Enrollment

- All student level variables were included in the model to determine those that are significant predictors of enrollment and their variation across schools

Table 6: Level 1 Logistic Coefficients Model Predicting College Enrollment

	Coefficient (SE)	Odds Ratio	t (df)	*p*
Intercept (γ00)	-1.32	0.27	-21.42 (378)	0.00
Gender_F (γ10)	0.23	1.25	11.50 (378)	0.00
Lunch (γ20)	-0.33	0.72	-14.309 (378)	0.00
HS_GPA (γ30)	0.61	1.85	16.192 (378)	0.00
10th_Reading (γ40)	0.001	1	3.572 (378)	0.00
SAT_Take (γ50)	0.91	2.49	30.03 (378)	0.00
PCT_HNRS (γ60)	1.01	2.75	8.15 (378)	0.00
Race_Black (γ70)	0.3	1.35	7.5 (378)	0.00
Race_His (γ80)	0.16	1.18	5.22 (378)	0.00
Race_Oth (γ90)	0.33	1.39	7.22 (378)	0.00
Random effects (var. components)	**Variance**	**df**	**Chi-Square**	***p***
Intercept (τ00)	1.11	278	1674.85	0.00

Results: Enrollment

- Lastly, all school-level variables were included in the model (significant interactions shown below)

Table 7: Contextual Logistic Model Predicting College Enrollment

	Coefficients (SE)	Odds Ratio	t (df)	p
Model for the intercepts				
Intercept (γ00)	-1.14 (.05)	0.32	-21.71 (244)	0.00
PCTABS21 (γ01)	-.03 (.01)	0.97	-3.20 (244)	0.002
PCT_ACT (γ04)	.01 (.004)	1.01	2.75 (244)	0.01
PCT_SAT (γ06)	-1.02 (.47)	0.36	-2.19(244)	0.03
PCT_PSAT (γ07)	.65 (.29)	1.92	2.20 (244)	0.03
PCT_Gift (γ010)	.05 (.01)	1.05	3.31 (244)	0.001
Num_FTE (γ015)	0.01 (.00)	1.01	3.95 (244)	0.00
Model for Gender_C slope				
Intercept (γ10)	.23 (.02)	1.25	9.05 (244)	0.00
PCT_Lunch (γ12)	.58 (.23)	1.78	2.55 (244)	0.01
Model for Lunch slope				
Intercept (γ20)	-0.33 (.03)	0.72	-11.65 (58,997)	0.00
TeachAvg (γ29)	-.02 (.01)	0.98	-2.21 (58,997)	0.03
Num_FTE (γ215)	-.002 (.001)	0.99	-2.89 (58,997)	0.00
Model for HSGPS slope				
Intercept (γ30)	.64 (.04)	1.89	17.33 (244)	0.00
PCT_Lunch (γ32)	-1.14 (.33)	0.32	-3.46 (244)	0.00
PCT_SAT (γ36)	-1.69 (.31)	0.18	-5.45 (244)	0.00
Num_FTE (γ315)	-.003 (.001)	0.99	-2.03 (244)	0.04
Model for 10th_Reading slope				
Intercept (γ40)	0.001 (.0004)	1.00	3.65 (58,997)	0.00
PCT_SAT (γ46)	-.01 (.003)	0.99	-2.25 (58,997)	0.02
Model for SAT_Take slope				
Intercept (γ50)	.87 (.03)	2.39	25.71 (244)	0.00
PCT_ACT (γ54)	-.01 (.002)	0.99	-2.79 (244)	0.01
PCT_SAT (γ56)	1.25 (.30)	3.51	4.16 (244)	0.00
Model for PCT_Honors slope				
Intercept (γ60)	.85 (.15)	2.34	5.81 (244)	0.00
Mean_SAT (γ65)	-.003 (.001)	1.0	-2.25 (244)	.03
PCT_Gift (γ610)	-.10 (.03)	.91	-2.85 (244)	.01
PCT_ELL (γ611)	-.08 (.03)	.92	-2.55 (244)	.01
Num_FTE (γ615)	-.01 (.01)	.99	-2.23 (244)	.03
Model for Race_BLA slope				
Intercept (γ70)	.23 (.04)	1.26	6.16 (58,997)	0.00
PCT_PSAT (γ77)	-.40 (.20)	.81	-2.02 (58,997)	.04
PCT_ELL (γ711)	-.02 (.01)	.98	-2.01 (58,997)	.04
DISC_TOT (γ712)	-.001 (.00)	1.0	-2.64 (58,997)	.01
Model for Race_HIS slope				
Intercept (γ80)	.11 (.05)	1.11	2.18 (58,997)	.03
Model for Race_OTH slope				
Intercept (γ90)	.36 (.06)	1.43	6.06 (58,997)	0.00
PCT_ELL (γ911)	.04 (.01)	1.04	2.98 (58,997)	0.00
NUM_FTE (γ915)	-.01 (.002)	.99	-2.96 (58,997)	.004

Random effects (var. components)	Variance	df	Chi-Square	p
Intercept (τ00)	.47	225	1025.08	*0.00*

Still a considerable amount of variance that could be further explained with additional measures.

Results: Persistence

Persistence =1, if a student with enroll = 1 had a record of enrollment in Fall 2002, Spring 2003, and Fall 2003 semesters.

Table 8: Empty Logistic Multilevel Model Predicting Persistence

Fixed Effects	Coefficient (SE)	Odds Ratio	t (df)	*p*
Intercept (γ_{00})	0.92 (0.03)	2.51	32.48 (362)	0.00

Random Effects	Variance	df	Chi square
τ_{00}	0.18	362	1436.20 (0.00)

- The probability of a student from a typical high school persisting through to the Fall 2003 academic term in a public in-state institution of higher education after enrolling in the Fall 2002 term is 71.5 percent (calculated as exp(.92)/(1 +exp(.92) = 2.51/3.51 = .715)
- The intraclass correlation coefficient for this model is .05 (calculated as ICC = τ00/(τ00+3.29) = .18/(.18+3.29) = .05)

Results: Persistence

- All student level variables were included in the model to determine those that are significant predictors of persistence and their variation across schools

Table 9: Level 1 Logistic Random Coefficients Model Predicting College Persistence

	Coefficient	Odds Ratio	t(df)	*p*
Intercept (γ00)	1.31	3.7	26.02 (335)	0.00
Gender (γ10)	0.09	1.1	2.31 (335)	0.02
LEP (γ20)	0.25	1.29	3.61 (335)	0.00
Lunch (γ30)	-0.3	0.74	-5.97 (335)	0.00
HS_GPA (γ40)	1.32	3.76	23.0 (335)	0.00
10th_Reading (γ50)	-0.002	1.00	-3.15 (335)	0.00
SATM (γ60)	0.001	1.00	3.29 (335)	0.00
PCT_HNRS (γ70)	1.03	2.79	6.14 (335)	0.00
AP (γ80)	0.35	1.42	7.08 (335)	0.00
Race_Bla (γ90)	0.3	1.34	5.17 (335)	0.00
Race_Oth (γ100)	0.48	1.62	5.29 (335)	0.00
Random effects (var. components)	**Variance**	**df**	**Chi-Square**	**p**
Intercept (τ00)	0.24	33	45.29	0.08

Results: Persistence

- Lastly, all school-level variables were included in the model (significant interactions shown below)

Table 10: Contextual Logistic Model Predicting College Persistence

	Coefficients (SE)	Odds Ratio	t (df)	p
Model for the intercepts				
Intercept (γ00)	1.22 (.06)	3.38	19.20 (226)	0.00
PCT_Gift (γ010)	.04 (.01)	1.05	3.09 (226)	0.00
Num_FTE (γ015)	.01 (.00)	1.01	2.61 (226)	0.01
Model for Gender_C slope				
Intercept (γ10)	.11 (.06)	1.12	1.87 (18,623)	0.06
Model for LEP slope				
Intercept (γ20)	-.01 (.17)	0.99	-.04 (18,623)	0.97
Model for Lunch slope				
Intercept (γ30)	-.25 (.07)	0.78	-3.47 (18,623)	0.00
PCT ELL (γ311)	.04 (.01)	1.04	3.20 (18,623)	0.00
StudTeach (γ314)	.05 (.03)	1.05	1.98 (18,623)	0.05
Model for HSGPA slope				
Intercept (γ40)	1.40 (.09)	4.05	15.94 (18,623)	0.00
Model for 10th_Reading slope				
Intercept (γ50)	-.00 (.00)	1.00	-1.00 (18,623)	0.32
Model for SAT_M slope				
Intercept (γ60)	.00 (.00)	1.00	1.30 (18,623)	0.20
PCT_ACT (γ64)	-.00 (.00)	1.00	-3.02 (18,623)	0.00
Model for PCT_Honors slope				
Intercept (γ70)	1.14 (.28)	3.12	4.00 (226)	0.00
PCT_ACT (γ74)	.04 (.02)	1.04	2.07 (226)	0.04
Model for AP slope				
Intercept (γ80)	.37 (.07)	1.44	4.99 (18,623)	0.00
Num_FTE (γ815)	-.01 (.00)	0.99	-2.48 (18,623)	0.01
Model for Race_BLA slope				
Intercept (γ90)	.40 (.10)	1.5	4.22 (18,623)	0.01
Model for Race_Other slope				
Intercept (γ100)	.58 (.15)	1.21	-.49 (18,623)	0.00

Random effects (var. components)	Variance	df	Chi-Square	p
Intercept (τ00)	.09	215	382.40	0.00

Persistence model warrants further investigation!

CollegeBoard inspiring minds

Results: Graduation

Graduate = 1, if a student with enroll = 1 and persist = 1 had obtained a Bachelor's degree from a public in-state university within five years from high school graduation.

Table 11: Empty Logistic Multilevel Model Predicting Graduation

Fixed Effects	Coefficient (SE)	Odds Ratio	t (df)	*p*
Intercept (γ_{00})	-0.48 (0.04)	0.62	-13.0 (353)	0.00

Random Effects	Variance	df	Chi square
τ_{00}	0.33	353	2021.47 (0.00)

- The probability of a student coming from a typical high school that enrolls in a public in-state higher education institution in Fall 2002 and persisting through to Fall 2003 academic term graduating with a Bachelor's degree within 5 academic years is 38.3 percent (calculated as exp(-.48)/(1 + exp(-.48) = 0.62/(1 + 0.62) = 0.383).
- The intraclass correlation coefficient for this model is .09

(calculated as ICC = τ00/(τ00+3.29) = .33/(.33+3.29) = .09)

Results: Graduation

- All student level variables were included in the model to determine those that are significant predictors of persistence and their variation across schools

Table 12: Logistic Random Coefficients Model Predicting College Graduation

	Coefficient	Odds Ratio	t(df)	*p*
Intercept (γ00)	-0.51 (.06)	0.6	-.8.58 (328)	0.00
Gender (γ10)	0.36 (.04)	1.44	8.62 (328)	0.00
Lunch (γ20)	-.32 (.04)	0.73	-6.21 (328)	0.00
HS_GPA (γ30)	2.0 (.06)	7.38	32.12 (328)	0.00
10th_Reading (γ40)	.00 (.00)	1.00	5.04 (328)	0.00
SATM (γ50)	.00 (.00)	1.00	5.03 (328)	0.00
PCT_HNRS (γ60)	2.37 (.17)	10.65	13.89 (328)	0.00
AP (γ70)	.22 (.05)	1.25	4.52 (328)	0.00
Race_Asi (γ80)	-.40 (.10)	0.67	-4.09 (328)	0.00
Race_Bla (γ90)	.20 (.07)	1.22	2.82 (328)	0.01
Race_His (γ100)	.38 (.07)	0.68	-5.29 (328)	0.00
Race_Oth (γ110)	.21 (.08)	1.23	2.63 (328)	0.01
Random effects (var. components)	**Variance**	**df**	**Chi-Square**	***p***
Intercept (τ00)	0.63	38	73.32	0.00

Results: Graduation

- Lastly, all school-level variables were included in the model (significant interactions shown below)

Table 13: Contextual Logistic Model Predicting College Graduation

Model for the intercepts	Coefficients (SE)	Odds Ratio	t (df)	*p*
Intercept (γ00)	-.67 (.07)	0.51	-9.43 (222)	0.00
PCT_Lunch (γ02)	-1.32 (.64)	0.27	-2.07 (222)	0.04
Mean_ACT (γ03)	.16 (.06)	1.18	2.63 (222)	0.01
PCT_ACT (γ04)	.01 (.01)	1.01	2.07 (222)	0.04
Model for Gender_C slope				
Intercept (γ10)	.47 (.06)	1.60	7.48 (15,004)	0.00
Model for Lunch slope				
Intercept (γ20)	-.46 (.08)	0.63	-5.61 (15,004)	0.00
PCTABS21 (γ21)	-0.02 (.01)	0.98	-2.03 (15,004)	0.04
Mean_SAT (γ25)	0.002 (.00)	1.00	2.01 (15,004)	0.04
PCT_PSAT (γ27)	.76 (.36)	2.15	2.12 (15,004)	0.03
StudTeach (γ214)	.08 (.03)	1.08	2.60 (15,004)	0.01
Model for HSGPA slope				
Intercept (γ30)	2.06 (.10)	7.86	19.90 (15,004)	0.00
Model for 10th_Reading slope				
Intercept (γ40)	.01 (.00)	1.01	4.23 (15,004)	0.00
PCTSAT (γ46)	-.03 (.01)	0.98	-2.67 (15,004)	0.01
Model for SAT_M slope				
Intercept (γ50)	.001 (.00)	1.00	3.79 (15,004)	0.00
PCT_Lunch (γ52)	-.01 (.00)	0.99	-2.02 (15,004)	0.04
TeachAvg (γ59)	-.0004 (.00)	1.00	-2.33 (15,004)	0.02
PCT_ELL (γ511)	.0002 (.00)	1.00	2.02 (15,004)	0.04
Model for PCT_Honors slope				
Intercept (γ60)	2.04 (.29)	7.70	7.10 (15,004)	0.00
PCTSAT (γ66)	5.06 (2.04)	158.11	2.48 (15,004)	0.01
TeachAvg (γ69)	.19 (.09)	1.21	2.09 (15,004)	0.04
StudTeach (γ614)	-.26 (.10)	0.77	-2.71 (15,004)	0.01
Model for AP slope				
Intercept (γ70)	.23 (.07)	1.26	3.34 (15,004)	0.00
PCT_Gift (γ710)	-.03 (.01)	0.97	-2.14 (15,004)	0.03
Model for Race_ASI slope				
Intercept (γ80)	-.49 (.16)	0.61	-2.98 (15,004)	0.00
PCTACT (γ84)	-.02 (.01)	0.98	-2.18 (15,004)	0.03
TeachAvg (γ89)	.12 (.05)	1.13	2.30 (15,004)	0.02
Model for Race_BLA slope				
Intercept (γ90)	.38 (.11)	1.47	3.53 (15,004)	0.00
PCTABS21 (γ91)	-.03 (.01)	0.97	-2.33 (15,004)	0.02
Model for Race_HIS slope	**Coefficients (SE)**	**Odds Ratio**	**t (df)**	**p**
Intercept (γ100)	-.35 (.14)	0.71	-2.46 (15,004)	0.01
PCTTeach (γ08)	-.03 (.01)	0.97	-2.16 (15,004)	0.03
TeachAvg (γ09)	.11 (.04)	1.12	3.15 (15,004)	0.00
Model for Race_Other slope				
Intercept (γ110)	.27 (.13)	1.31	2.14 (15,004)	0.03

Random effects (var. components)	Variance	df	Chi-Square	*p*
Intercept (τ00)	.09	222	435.28	0.00

Discussion

- Student- and school-level characteristics are important predictors of educational outcomes!
- Results consistent with the College Board's notion of College Readiness
  - Multiple measures (academic & non-cognitive) should be taken into account when assessing whether students are ready for college.
  - Academic measure = HSGPA + SAT + Academic Rigor
- Evidence that AP test-taking is linked with success in college.

Limitations

- Data Source
  - Not all institutions provide data (or accurate data) to The National Student Clearinghouse
- Definition of Persistence
  - Too stringent?
- Estimation method
  - Laplace vs restricted PQL

Future Research

- Replicate study using the College Board's new College Readiness indicator (HSGPA, composite SAT Scores and the new academic rigor index)

- Conduct 3 level HGLM's to examine how characteristics of institutions of higher education play a role in students' decision to enroll.

Questions and Comments

- Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board presentations do not necessarily represent official College Board position or policy.
- Questions should be directed to kgodfrey@collegeboard.org & hmatoselefonte@collegeboard.org

CollegeBoard
inspiring minds