

Early Indicators of Student Success: A Multi-State Analysis

Abstract

This paper reports the results of a four-state collaboration that uses Student Unit Record Systems that track students from high school into college. The goal is to determine whether it is possible to accurately predict or identify which students will not graduate using very early indicators – variables available at college entry or during the first semester. Using statistical models we are able to identify those students at greatest risk of non-completion quite accurately at this very early stage, allowing college staff to prioritize interventions and supports aimed at improving retention and completion for those at greatest risk. Our models do not use gender, race or ethnicity in determining probability of non-completion.

Introduction

In recent years, several states have compiled databases that track multiple cohorts of students from school into college and later into the labor force. Known as Student Unit Record Systems (SURS), these typically compile socio-demographic information for hundreds of thousands of high school students along with measures of academic performance such as scores on state tests and high school GPA. For those who continue into college, SURS compile information on colleges attended and transcript details of grades and credits and major field of study. Some states have also incorporated students' quarterly earnings before, during and after college, drawn from state departments of labor. Dynarski and Berends (2015) argue that SURS are a game changer for educational research.

In this paper, we present collaborative research analyzing SURS data from four different states – Texas, New York, Virginia and Illinois – to examine degree completion. By studying four state SURS individually, each using data on past cohorts of students who had been tracked for sufficient years to graduate, we determine whether the same statistical model can predict graduation in different states, assessing the generalizability of findings.

Our central research question is whether it is possible to use SURS data to predict accurately which undergraduates did not graduate. A large body of research (summarized below) has identified factors that *on average* are associated with degree non-completion. The use of SURS in this paper has a different goal: to predict *for individual students* who are likely to complete the credential. If one can determine with a reasonable degree of accuracy from the first semester of college or even before entry to college which individuals are at high risk of non-completion, this information can be used to reach out to those particular students, to provide counseling or support, or to intervene in other ways, in order to increase retention and graduation.

Several commercial information systems sold to colleges (e.g., Starfish, EAB, Civitas) already provide early warnings or alerts that a particular student is at risk of dropping out or failing a course. These systems typically require faculty members to input information for their classes about students' weekly attendance and grades on midterms and assignments. Based on analyses of these data, the programs issue alerts that “flag” or identify students at high risk of failure in that particular course, allowing college staff to intervene to support to students in difficulty.

By contrast, SURS do not require special data collection efforts by faculty for their classes: they use data already known for each student at entry to college, along with information collected at registration on course-load and remedial course placements. Later models include additional measures of student performance in the first semester of college.

This paper finds that it is possible to accurately predict individuals' probability of graduating, using only very early SURS indicators. This contrasts with most of the commercial alert systems that focus on predicting passing or failure for a particular course. Using SURS data, we generate predictions using conventional statistical models enhanced with some data mining techniques. We find that a 'standard' model yields accurate predictions of non-completion across four different states and is most accurate for those students at highest risk of non-completion.

Prior Literature

Tinto (1988, 1994, 2012) developed a theory emphasizing the importance of a student's academic and social integration for persistence in college, arguing that a lack of fit between a student and the college was a proximal cause of dropping out. Using national survey data, Clifford Adelman (1999, 2006) developed a theory of academic momentum, in which student progress in the first year of college, specifically completing 20 or more credits, was predictive of degree completion. Students who completed fewer credits were less likely to persist. Adelman (1999) identified academic preparation as central to sustaining momentum: students who did not take a rigorous curriculum during high school (most especially in mathematics) and consequently faced difficulties in college were at high risk of non-completion (cf. Chingos 2018). The idea that passing college mathematics courses, especially remedial math, constitute a major hurdle to degree completion has led to widespread efforts to reform that part of the curriculum (Bailey, Jeong, & Cho 2010; Chen & Simone 2016; Hayward & Willett 2014).

Other research emphasizes competing demands faced by many undergraduates, who need to juggle academic studies with paid employment and family obligations, creating time binds that lower graduation rates (Stinebrickner & Stinebrickner 2003, 2004; Bozick 2007). St. John (2003), Goldrick-Rab (2016) and others also highlight the role of finances, arguing that inadequate financial aid generates financial stresses leading some students to drop out because they cannot afford to continue.

This literature has identified important factors associated with retention/attrition and completion, and has estimated average effects of predictors across representative samples of students. It has not, however, focused on predicting completion outcomes for individual students, the goal of this paper.

Methods

Researchers in each state had access to and analyzed their own state's SURS database. In three states, data were analyzed separately for public four-year colleges and for two-year community colleges. For one state (Illinois) only data on community colleges were available. Although data on private college enrollments were available in some SURS, private colleges did not usually

report transcript information, so they were excluded from the analyses for this paper. The SURS data were longitudinal, following each student from entry into college for at least ten years, at which time the SURS indicated whether that individual had graduated or not.

Using logistic regression and some exploratory data mining methods we built models to predict graduation, either at that institution or at another, by the SURS end date. Given the large number of students who transfer from two-year to four-year colleges, many of whom do not obtain an associate's degree before transferring, the 'graduation' milestone for students who began at two-year colleges was defined as either completion of an associate's or a bachelor's degree or accumulation of sixty credits (the latter approximating AA completion).

After exploratory modeling we settled on a logistic regression containing the following independent variables for analyses of retention into the second year: age at college entry, parental adjusted gross income ("AGI"), high school GPA, SAT score (if available), remedial requirements (math, reading, writing), workload in semester 1 (total number of credits counting both remedial and non-remedial courses), and whether a major was declared at college entry (a dichotomous variable). We converted all continuous variables into categorical predictors; this also allowed us to add an additional category for "missing" for each variable.

Several additional independent variables captured student progress in the first semester: (non-remedial) course credits earned in semester 1, average (non-remedial course) GPA in semester 1, and remedial math/reading/writing performance (if remediation was required). Appendix A reports details for each variable.

Readers should note that we deliberately avoided using gender, race or ethnicity as predictors of retention and graduation in statistical models. We avoided using group identities as predictors from a concern that incorporating such attributes in our predictive algorithm could lead to forms of group-based stereotyping or statistical discrimination. After completing our analyses without such attributes, we checked (separately for each state SURS) to see whether the inclusion of gender and race/ethnicity would have improved predictive accuracy. Adding those variables made very little if any improvement in predictive accuracy, given the behavioral measures already in the model, so we did not lose accuracy by omitting those factors.

There were some differences in data availability between states. For example, Texas used a statewide assessment of math and English skills that is mandatory for high school seniors and transformed this into a percentile score, whereas the New York data used SAT scores. Different state SURS also had somewhat different measures of low income: some used eligibility for free or subsidized school lunches in high school, while others used Pell eligibility or adjusted parental income. Consequently, the variables in models are not identical across the states, though they are quite similar.

Findings

For each state, a logistic model was run predicting graduation, separately for community college and four-year college entrants. Appendix B reports those logistic models and their coefficients. Generally, the most powerful predictors for graduation were consistent with the academic

momentum perspective: credits earned in the first semester and first semester GPA were the strongest predictors. Among community college entrants, the next most powerful predictor was each student's status regarding remedial math – whether the student was required to take remedial math, and if so whether the student passed or withdrew/failed, or whether the student avoided taking the remedial course in the first semester. For students entering four-year colleges, the strongest predictors were again academic momentum in the first semester. However, high school GPA, adjusted parental income, and age at entry were also important factors associated with graduation.

For each state, the same logistic model yielded a predicted probability of graduation for each student. We grouped those probabilities into percentiles and then deciles. For each decile, we could also measure what percentage of students in that decile actually graduated, because the SURS data contained the actual outcome. This yields a measure of the accuracy of prediction for each part of the probability distribution.

Table 1 and 2 report these statistics for each state and separately for two-year and four-year entrants. The resulting pattern is in every case curvilinear. The predictive accuracy of the model is very high for students least likely to graduate and is quite high for students most likely to graduate. For students with predicted probabilities in the mid-range, however, the accuracy of the model is much lower: students some middle deciles have closer a 50:50 change of graduating, for example. The model is far less accurate for this middle group of students.

If the goal were accurate prediction for every student across the distribution, this would be a serious drawback. However, if the goal is to provide actionable information identifying those students highest risk of not graduating, the model successfully identifies those students. In marketing and data mining terminology, the model has high “lift” for students in the deciles at the extremes.

Discussion and Conclusion

We determined that it is possible, using SURS data, to identify those students who are least likely to graduate, using one very similar model across states. The predictive accuracy of the model has a curvilinear shape for all four state SURS: the highest predictive accuracy – usually exceeding 90 percent -- occurs for students at highest risk of non-completion. Prediction in the middle of the distribution is much less accurate.

Using early indicators and correctly predicting 8 or 9 times out of ten those at greatest risk enables college staff to target interventions or support services, without the arduous data entry burden of commercial early alert systems. Colleges can prioritize or customize approaches to those most or least likely to complete a degree using this model. At these high levels of accuracy, a small proportion of students will nevertheless be mistakenly identified as at risk (false positives), but so long as that determination only results in additional attention and support, little harm is done. We conclude that SURS data can play a useful role in efforts to increase undergraduate retention and graduation, without the large investment and faculty effort required for commercial alert systems.

References

- Adelman, Clifford. 1999. *Answers in the Toolbox: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment*. Available at <http://www.ed.gov/pubs/Toolbox/toolbox.html>
- Adelman, Clifford. 2006. *The Toolbox Revisited: Paths to Degree Completion from High School Through College*. Washington DC: U.S. Department of Education. Available at
- Bailey, Thomas, Dong Wook Jeong, & Sung-Woo Cho. (2010). "Referral, enrollment, and completion in developmental education sequences in community colleges." *Economics of Education Review*, 29(2):255-270.
- Bozick, Robert. 2007. "Making it through the first year of college: The role of students' economic resources, employment, and living arrangements." *Sociology of Education*, 80:261-285.
- Chen, Xianglei, & Sean Simone. 2016. *Remedial Course Taking at U.S. Public 2- and 4-Year Institutions: Scope, Experiences, and Outcomes* (NCES 2016-405). Washington DC: U.S. Department of Education, National Center for Education Statistics.
- Chingos, Matthew. 2018. *What Matters Most for College Completion? Academic Preparation is the Key Predictor of Success*. Washington DC: American Enterprise Institute (AEI).
- Dynarski, Susan, & Mark Berends. 2015. "Introduction to the Special Issue." *Education Evaluation and Policy Analysis*, 37(1):3s-5s.
- Goldrick-Rab, Sara. 2016. *Paying the Price: College Costs, Financial Aid, and the Betrayal of the American Dream*. Chicago: University of Chicago Press.
- Hayward, Craig, & Terrence Willett. 2014. *Curricular Redesign and Gatekeeper Completion: A Multi-College Evaluation of the California Acceleration Project*. San Rafael, CA: The RP Group.
- Stinebrickner, Ralph, & Todd R. Stinebrickner. 2003. "Working during school and academic performance." *Journal of Labor Economics*, 21(2):473-491.
- Stinebrickner, Ralph, & Todd R. Stinebrickner. 2004. "Time-use and college outcomes." *Journal of Econometrics*, 121(1/2):243-269.
- St. John, Edward P. 2003. *Refinancing the College Dream: Access, Equal Opportunity, and Justice for Taxpayers*. Baltimore: Johns Hopkins University Press.
- Tinto, Vincent. 1988. "Stages of student departure: Reflections on the longitudinal character of student leaving." *The Journal of Higher Education*, 59(4):438-455.

Tinto, Vincent. 1994. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: University of Chicago Press.

Tinto, Vincent. 2012. *Completing College: Rethinking Institutional Action*. Chicago: University of Chicago Press.

Table 1: AA students' predicted probability distribution of graduation by actual % of students that did not graduate

Probability group by model prediction on test data	New York (N=107,351)		Texas (N=46,244)		Virginia (N=60,085)		Illinois (N=34,828)	
	Model A ^a	Model B ^b	Model A	Model B	Model A	Model B	Model A	Model B
bottom 1% (least likely to graduate)	86.90	97.95	94.34	98.49	85.00	99.18	93.14	97.44
bottom 5%	84.45	96.52	91.75	96.20	87.41	97.63	86.94	97.36
bottom 10%	83.29	95.46	90.64	95.63	87.77	96.98	85.79	97.50
2nd decile	79.27	90.90	87.55	92.69	87.91	94.40	81.07	95.64
3rd decile	74.67	85.10	84.01	88.95	83.83	91.97	76.22	89.58
4th decile	72.13	78.32	82.64	83.95	82.88	86.74	70.87	82.03
5th decile	69.65	71.36	78.06	78.14	78.21	83.08	68.08	71.11
6th decile	64.73	63.13	73.01	71.96	76.85	76.62	63.04	63.16
7th decile	59.87	55.64	65.36	59.45	69.92	69.57	58.61	51.94
8th decile	55.67	46.22	47.04	45.21	64.10	60.35	53.19	41.99
9th decile	48.25	35.58	26.71	24.72	57.03	47.12	45.16	30.20
10th decile (most likely to graduate)	35.92	21.80	15.23	10.08	44.41	28.87	34.28	15.42
% did not graduate overall	64.35	64.35	65.08	65.08	73.58	73.58	63.86	63.86

Note: ^a Variables at college entry. ^b Variables at college entry + 1st semester variables.

Table 2: BA students' predicted probability distribution of graduation by actual % of students that did not graduate

Probability group by model prediction on test data	New York (N=47,045)		Texas (N=20,736)		Virginia (N=63,773)	
	Model A ^a	Model B ^b	Model A	Model B	Model A	Model B
bottom 1% (least likely to graduate)	87.42	96.96	91.83	98.56	76.18	94.98
bottom 5%	81.42	95.66	87.15	96.53	65.07	86.58
bottom 10%	76.22	92.90	82.74	94.17	60.33	78.07
2nd decile	67.12	80.22	68.38	81.73	48.74	52.96
3rd decile	61.29	67.59	60.42	66.57	38.47	38.92
4th decile	58.12	59.54	53.24	54.58	30.25	27.94
5th decile	52.61	50.56	47.58	45.24	24.89	20.57
6th decile	49.52	44.62	41.91	35.47	18.82	15.56
7th decile	46.84	38.83	34.00	28.96	16.78	12.06
8th decile	41.79	34.09	26.20	20.09	13.42	10.12
9th decile	35.06	26.20	21.52	15.25	9.14	7.39
10th decile (most likely to graduate)	21.47	15.90	13.66	10.44	7.04	4.50
% did not graduate overall	51.05	51.05	45.25	45.25	26.81	26.81

Note: ^a Variables at college entry. ^b Variables at college entry + 1st semester variables.

Appendix A: Variables descriptions (New York)

Gender

male
female

Race

white
black
Hispanic
Asian
native American

High school GPA

A-/A: 3.67-4.00
B+: 3.33-3.67
B: 3.00-3.33
B-: 2.67-3.00
C+: 2.33-2.67
no HS GPA record

SAT score

1st quintile
2nd quintile
3rd quintile
4th quintile
5th quintile
no SAT score

Age at entry

18 or younger
19
20
21
22
23
24
25 or older

Parental adjusted gross income

1st quartile (highest)
2nd quartile
3rd quartile
4th quartile
parent AGI missing
parent AGI missing for cohort

Major in semester 1

declared
not declared
unclassified (unknown)

Remedial requirements at entry

no remedial requirement
remedial math required only
remedial reading required only
remedial writing required only
two or more remedial requirements
remedial requirement unknown

Remedial math semester 1

not required, not taken
required, not taken
passed all
failed/withdrew one or more

Remedial reading semester 1

not required, not taken
required, not taken
passed all
failed/withdrew one or more

Remedial writing semester 1

not required, not taken
required, not taken
passed all
failed/withdrew one or more

Workload semester 1

< 8 credits
 ≥ 8 credits & < 12 credits
 ≥ 12 credits & < 14 credits
 ≥ 14 credits & < 16 credits
 ≥ 16 credits & < 18 credits
 ≥ 18 credits & < 20 credits
> 20 credits

Credits earned semester 1

0 credits (but enrolled)
> 0 credits & < 4 credits
 ≥ 4 credits & < 8 credits
 ≥ 8 credits & < 12 credits
 ≥ 12 credits & < 14 credits
 ≥ 14 credits & < 16 credits
 ≥ 16 credits & < 18 credits

≥ 18 credits & < 20 credits
20 credits or more

GPA semester 1 (non-remedial)

A-/A: 3.67-4.00

B+: 3.33-3.67

B: 3.00-3.33

B-: 2.67-3.00

C+: 2.33-2.67

C: 2.00-2.33

C-: 1.67-2.00

D+: 1.33-1.67

D: 1.00-1.33

D-/F: < 1.00

enrolled, no GPA record

Dependent variables

enrolled in 2nd year

graduated

Appendix B1: Predicting graduation with variables at entry and first semester (New York)

Variables	AA at entry Coefficients (SE)		BA at entry Coefficients (SE)	
	Entry variables	Entry + semester 1 variables	Entry variables	Entry + semester 1 variables
SAT quintiles (ref. = 3rd quintile)				
1st quintile (lowest)	-0.36*** (0.016)	0.052** (0.018)	-0.16*** (0.027)	-0.018 (0.029)
2nd quintile	-0.11*** (0.016)	0.071*** (0.018)	-0.032 (0.021)	0.031 (0.023)
4th quintile	0.019 (0.022)	-0.12*** (0.024)	0.028 (0.017)	0.0064 (0.018)
5th quintile (highest)	-0.074* (0.033)	-0.33*** (0.037)	0.15*** (0.017)	-0.0039 (0.018)
No SAT score	-0.62*** (0.014)	-0.27*** (0.016)	-0.20*** (0.028)	-0.12*** (0.030)
Workload semester 1 (ref. 12 to 13 credits)				
Less than 8 credits	-0.63*** (0.019)	-0.24*** (0.021)	-1.10*** (0.047)	-0.30*** (0.054)
8 to 11 credits	-0.38*** (0.016)	-0.14*** (0.017)	-0.57*** (0.042)	-0.20*** (0.045)
14 to 15 credits	0.16*** (0.0093)	0.10*** (0.011)	0.23*** (0.012)	-0.0065 (0.017)
16 to 17 credits	0.26*** (0.012)	0.088*** (0.014)	0.33*** (0.016)	-0.0042 (0.022)
18 to 19 credits	0.39*** (0.015)	0.12*** (0.017)	0.22*** (0.049)	-0.10 (0.059)
20 or more credits	0.73*** (0.016)	0.29*** (0.018)	0.071 (0.095)	0.059 (0.10)
Major selection at entry (ref. = declared major)				
Did not declare major	0.11** (0.039)	0.27*** (0.044)	0.032** (0.012)	0.011 (0.013)
Unclassified	0.032 (0.017)	-0.10*** (0.019)	0.33 (0.17)	0.23 (0.18)
HS GPA (ref. = C+ or less: 0-2.66)				
B-: 2.67-2.99	0.48*** (0.0098)	0.22*** (0.011)	0.50*** (0.026)	0.38*** (0.028)
B: 3.00-3.32	0.91*** (0.012)	0.45*** (0.014)	1.01*** (0.025)	0.72*** (0.027)
B+: 3.33 - 3.66	1.30*** (0.021)	0.69*** (0.023)	1.64*** (0.027)	1.04*** (0.029)

A-/A: 3.67-4.00	1.64*** (0.046)	0.91*** (0.051)	2.11*** (0.037)	1.24*** (0.040)
No HS GPA record	0.21*** (0.010)	0.11*** (0.011)	0.88*** (0.037)	0.48*** (0.040)
Age at entry (ref. = 19)				
18 or younger	0.21*** (0.012)	0.16*** (0.013)	0.10*** (0.013)	0.081*** (0.014)
20	-0.29*** (0.011)	-0.26*** (0.012)	-0.27*** (0.019)	-0.24*** (0.021)
21	-0.42*** (0.014)	-0.42*** (0.016)	-0.40*** (0.033)	-0.34*** (0.035)
22	-0.34*** (0.018)	-0.40*** (0.020)	-0.42*** (0.045)	-0.40*** (0.048)
23	-0.24*** (0.017)	-0.40*** (0.019)	-0.39*** (0.048)	-0.58*** (0.051)
24 or older	-0.021 (0.013)	-0.35*** (0.014)	-0.30*** (0.039)	-0.64*** (0.042)
Entry age missing	-2.84** (1.02)	-2.61* (1.04)	-1.39 (0.88)	-1.50 (0.98)
Parental adjusted gross income (ref. = bottom quartile)				
Parent AGI missing	-0.11*** (0.014)	-0.20*** (0.016)	-0.21*** (0.023)	-0.24*** (0.025)
2nd quartile	-0.015 (0.017)	-0.024 (0.019)	0.023 (0.025)	-0.011 (0.026)
3rd quartile	-0.16*** (0.017)	-0.20*** (0.019)	-0.030 (0.025)	-0.091*** (0.027)
Top quartile	-0.091*** (0.018)	-0.22*** (0.020)	0.11*** (0.023)	-0.0030 (0.025)
Parental AGI missing for cohort	-0.047*** (0.013)	-0.14*** (0.015)	0.21*** (0.020)	0.32*** (0.021)
Remedial requirement at entry (ref. = no remedial requirement)				
Remedial math required only	-0.073*** (0.011)			
Remedial reading required only	0.49*** (0.033)			
Remedial writing required only	0.28*** (0.013)			
2 or more remedial requirements	-0.21*** (0.0100)			
Remedial requirement unknown	0.028* (0.013)			
GPA semester 1 (ref. = D-/F: <1.00)				

D: 1.00-1.32	0.28*** (0.030)	0.28*** (0.077)
D+: 1.33-1.66	0.31*** (0.029)	0.53*** (0.070)
C-: 1.67-1.99	0.47*** (0.029)	0.77*** (0.068)
C: 2.00-2.32	0.70*** (0.026)	0.96*** (0.066)
C+: 2.33-2.66	0.86*** (0.027)	1.20*** (0.066)
B-: 2.67-2.99	1.03*** (0.026)	1.39*** (0.066)
B: 3.00-3.32	1.25*** (0.025)	1.63*** (0.065)
B+: 3.33 - 3.66	1.53*** (0.028)	1.92*** (0.066)
A-/A: 3.67-4.00	1.71*** (0.026)	2.20*** (0.066)
No GPA record	0.99*** (0.024)	1.08*** (0.085)
Credits earned semester 1 (ref. = 12 to 13 credits)		
0 credits	-1.79*** (0.034)	-1.66*** (0.090)
1 to 3 credits	-1.31*** (0.019)	-1.27*** (0.053)
4 to 7 credits	-0.83*** (0.016)	-0.83*** (0.027)
8 to 11 credits	-0.38*** (0.015)	-0.38*** (0.017)
14 to 15 credits	0.24*** (0.024)	0.31*** (0.020)
16 to 17 credits	0.45*** (0.039)	0.52*** (0.029)
18 to 19 credits	0.72*** (0.061)	0.39** (0.12)
20 or more credits	0.85*** (0.089)	0.69 (0.54)
Remedial math semester 1 (ref. = not required, not taken)		
Required, not taken	-0.44*** (0.016)	
Passed all	0.029** (0.010)	
Failed/withdrew one or more	-0.54***	

		(0.013)		
Remedial reading semester 1 (ref. = not required, not taken)				
Required, not taken		0.13***		
		(0.022)		
Passed all		0.27***		
		(0.013)		
Failed/withdrew one or more		-0.30***		
		(0.023)		
Remedial writing semester 1 (ref. = not required, not taken)				
Required, not taken		0.17***		
		(0.014)		
Passed all		0.14***		
		(0.011)		
Failed/withdrew one or more		-0.26***		
		(0.015)		
Constant	-0.36***	-0.33***	-1.25***	-2.01***
	(0.019)	(0.033)	(0.033)	(0.073)
Observations	357,836	357,836	153,795	153,795

Standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix B2: Predicting graduation with variables at entry and first semester (Texas)

Variables	AA at entry Odds Ratios (SE)		BA at entry Odds Ratios (SE)	
	Entry variables	Entry + semester 1 variables	Entry variables	Entry + semester 1 variables
TAKS-reading quintiles (ref. = 3rd quintile)				
1st quintile	0.858*** (0.030)	0.902** (0.033)		
2nd quintile	0.950 (0.031)	0.968 (0.033)		
4th quintile	1.089* (0.040)	1.048 (0.041)		
5th quintile (highest)	1.158*** (0.049)	1.067 (0.048)		
No TAKS-reading score	0.927 (0.070)	0.849* (0.070)		
TAKS-math quintiles (ref. = 1st quintile)				
1st quintile	0.706*** (0.024)	0.807*** (0.029)		
2nd quintile	0.901* (0.029)	0.945 (0.032)		
4th quintile	1.186*** (0.042)	1.107** (0.041)		
5th quintile (highest)	1.349*** (0.060)	1.153** (0.054)		
No TAKS-math score	0.967 (0.074)	0.969 (0.080)		
SAT quintiles (ref. = 1st quintile)				
2nd quintile			1.368*** (0.053)	1.198*** (0.051)
3rd quintile			1.646*** (0.066)	1.252*** (0.057)
4th quintile			1.863*** (0.075)	1.284*** (0.059)
5th quintile (highest)			2.408*** (0.103)	1.382*** (0.067)
No SAT score			0.902* (0.045)	0.689*** (0.039)
Workload semester 1 (ref. 12 to 13 credits)				
Less than 8 credits	0.631***	0.951	0.790***	1.281**

	(0.013)	(0.028)	(0.045)	(0.105)
8 to 11 credits	0.871***	1.008	0.575***	0.766***
	(0.019)	(0.024)	(0.039)	(0.060)
14 to 15 credits	1.121***	0.984	1.307***	0.977
	(0.031)	(0.032)	(0.030)	(0.036)
16 to 17 credits	1.138***	0.925	1.436***	0.852**
	(0.043)	(0.042)	(0.044)	(0.044)
18 to 19 credits	1.279***	1.015	1.607***	0.790
	(0.058)	(0.054)	(0.148)	(0.105)
20 or more credits	1.317***	1.052	2.054***	0.781
	(0.049)	(0.057)	(0.240)	(0.161)
Major selection at entry (ref. = Declared major)				
Did not declare major	1.052***	1.077***	0.809***	0.865***
	(0.016)	(0.017)	(0.018)	(0.021)
HS class rank by quartile (ref. = 3rd quartile)				
1 st quartile (highest)	0.606***	0.562***	3.200***	2.345***
	(0.041)	(0.038)	(0.088)	(0.071)
2 nd quartile	0.777***	0.746***	1.922***	1.568***
	(0.041)	(0.039)	(0.047)	(0.042)
4 th quartile (lowest)	0.608***	0.742***	0.790**	0.846*
	(0.012)	(0.015)	(0.061)	(0.072)
Age at entry (ref. = 19)				
18 or younger	1.154***	1.126***	1.171***	1.153***
	(0.028)	(0.028)	(0.043)	(0.047)
20	0.956	0.977	0.601***	0.648**
	(0.040)	(0.041)	(0.076)	(0.090)
21	0.920	0.917	0.672*	0.716
	(0.046)	(0.047)	(0.115)	(0.134)
22	1.078	1.044	1.140	1.053
	(0.057)	(0.056)	(0.218)	(0.222)
23	1.075	1.013	1.151	0.961
	(0.062)	(0.060)	(0.237)	(0.219)
24	1.051	0.974	1.512	1.214
	(0.063)	(0.059)	(0.355)	(0.324)
25 or older	1.290***	1.148***	1.018	0.813
	(0.041)	(0.038)	(0.115)	(0.101)
Parental adjusted gross income (ref. = bottom quartile)				
Parent AGI missing	1.024	0.944*	1.301***	0.966
	(0.024)	(0.023)	(0.069)	(0.056)
2nd quartile	1.116***	1.051	1.179***	1.163***
	(0.028)	(0.027)	(0.047)	(0.051)
3rd quartile	1.177***	1.056*	1.354***	1.292***

	(0.031)	(0.028)	(0.051)	(0.053)
Top quartile	0.987	0.866***	2.085***	1.853***
	(0.029)	(0.025)	(0.072)	(0.070)
Remedial requirement at entry (ref. = no remedial requirement)				
Remedial math required only	0.901***	0.993	0.662***	0.818**
	(0.019)	(0.028)	(0.035)	(0.055)
Remedial reading required only	1.046	1.109	0.761**	0.801
	(0.049)	(0.062)	(0.072)	(0.106)
Remedial writing required only	0.947	1.082	0.642***	0.687**
	(0.052)	(0.067)	(0.063)	(0.092)
2 or more remedial requirements	0.646***	0.903**	0.536***	0.598***
	(0.013)	(0.036)	(0.024)	(0.073)
GPA semester 1 (ref. = D-/F: <1.00)				
D: 1.00-1.32		1.640***		2.647***
		(0.080)		(0.220)
D+: 1.33-1.66		1.983***		3.764***
		(0.094)		(0.281)
C-: 1.67-1.99		2.429***		4.892***
		(0.149)		(0.364)
C: 2.00-2.32		2.407***		7.175***
		(0.092)		(0.481)
C+: 2.33-2.66		2.756***		10.175***
		(0.108)		(0.675)
B-: 2.67-2.99		2.945***		12.651***
		(0.146)		(0.864)
B: 3.00-3.32		2.880***		16.925***
		(0.102)		(1.112)
B+: 3.33 - 3.66		3.010***		21.728***
		(0.119)		(1.476)
A+/A: 3.67-4.00		3.133***		25.308***
		(0.114)		(1.760)
Credits earned semester 1 (ref. = 4 to 7 credits)				
0 credits		1.336***		2.174***
		(0.055)		(0.387)
1 to 3 credits		0.733***		0.515***
		(0.020)		(0.054)
8 to 11 credits		1.290***		1.842***
		(0.034)		(0.119)
12 to 13 credits		1.515***		2.802***
		(0.050)		(0.185)
14 to 15 credits		1.540***		3.404***
		(0.077)		(0.253)

16 to 17 credits		1.474*** (0.099)		3.889*** (0.346)
18 to 19 credits		1.492*** (0.116)		4.469*** (0.839)
20 or more credits		1.347*** (0.102)		5.381*** (1.352)
Remedial math semester 1 (ref. = not required, not taken)				
Required, not taken		0.896*** (0.029)		0.931 (0.072)
Passed all		1.297*** (0.032)		1.108* (0.050)
Failed/withdrew one or more		0.775*** (0.026)		0.669*** (0.042)
Remedial reading semester 1 (ref. = not required, not taken)				
Required, not taken		0.960 (0.038)		1.105 (0.128)
Passed all		1.250*** (0.041)		1.175 (0.106)
Failed/withdrew one or more		0.612*** (0.036)		0.584*** (0.102)
Remedial writing semester 1 (ref. = not required, not taken)				
Required, not taken		0.877*** (0.032)		1.104 (0.124)
Passed all		1.230*** (0.037)		1.288** (0.112)
Failed/withdrew one or more		0.677*** (0.037)		0.780 (0.118)
Constant	0.443*** (0.020)	0.141*** (0.008)	0.276*** (0.015)	0.021*** (0.002)
Observations	107,983	107,983	48,783	48,783

Standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix B3: Predicting graduation with variables at entry and first semester (Virginia)

Variables	AA at entry Odds Ratios (SE)		BA at entry Odds Ratios (SE)	
	Entry variables	Entry + semester 1 variables	Entry variables	Entry + semester 1 variables
SAT quintiles (ref. = 1st quintile)				
2nd quintile (2nd lowest)	0.866 (0.112)	0.915 (0.131)	1.121*** (0.017)	1.142*** (0.019)
3rd quintile	0.733 (0.127)	0.791 (0.152)	1.252*** (0.020)	1.228*** (0.021)
4th quintile	0.481** (0.118)	0.506* (0.136)	1.389*** (0.024)	1.280*** (0.024)
5th quintile (highest)	1.506 (0.690)	1.681 (0.874)	1.947*** (0.039)	1.479*** (0.031)
No SAT score	0.752*** (0.043)	0.564*** (0.035)	0.816*** (0.014)	0.829*** (0.015)
Workload semester 1 (ref. = less than 8 credits)				
8 to 11 credits	1.864*** (0.024)	1.112*** (0.023)	1.535*** (0.108)	0.891 (0.067)
12 to 13 credits	2.977*** (0.044)	1.296*** (0.034)	2.457 (0.157)	0.883 (0.062)
14 to 15 credits	4.659*** (0.085)	1.583*** (0.051)	3.657*** (0.233)	0.947 (0.066)
16 to 17 credits	5.803*** (0.139)	1.441*** (0.063)	3.965*** (0.253)	0.830** (0.059)
18 to 19 credits	1.127*** (0.094)	1.515*** (0.129)	3.617*** (0.257)	0.617*** (0.053)
20 or more credits	6.769*** (0.350)	0.884 (0.220)	2.939*** (0.352)	0.781 (0.135)
Missing	0.766*** (0.018)	0.517*** (0.025)	2.522*** (0.240)	2.316*** (0.235)
Major selection at entry (ref. = declared major)				
Did not declare major	2.472*** (0.350)	2.834*** (0.428)	1.017 (0.010)	1.050*** (0.011)
HS GPA (ref. = C+)				
B-	1.376*** (0.075)	1.182** (0.069)	1.346*** (0.033)	1.191*** (0.032)
B	1.987*** (0.112)	1.396*** (0.085)	2.127*** (0.049)	1.620*** (0.041)
B+	2.965***	1.743***	3.394***	2.156***

	(0.200)	(0.127)	(0.080)	(0.055)
A/A+	4.389***	2.031***	6.810***	3.344***
	(0.380)	(0.188)	(0.163)	(0.088)
Missing	1.244***	1.147***	2.793***	1.748***
	(0.049)	(0.049)	(0.069)	(0.047)
Age at entry (ref. = 18 or younger)				
19	0.800***	0.804***	0.938***	0.933***
	(0.010)	(0.011)	(0.012)	(0.013)
20	0.663***	0.631***	0.593***	0.587***
	(0.014)	(0.015)	(0.026)	(0.028)
21	0.626***	0.546***	0.546***	0.521***
	(0.018)	(0.017)	(0.041)	(0.043)
22	0.655***	0.518***	0.415***	0.378***
	(0.021)	(0.018)	(0.040)	(0.039)
23	0.691***	0.525***	0.525***	0.453***
	(0.024)	(0.019)	(0.061)	(0.056)
24	0.667***	0.478***	0.447***	0.392***
	(0.024)	(0.018)	(0.058)	(0.055)
25 or older	0.721***	0.488*	0.523***	0.425***
	(0.011)	(0.008)	(0.037)	(0.032)
Parental adjusted gross income (ref. = bottom quartile)				
Parent AGI missing	1.389***	1.235***	1.879***	1.735***
	(0.019)	(0.018)	(0.034)	(0.034)
2nd quartile	1.432***	1.223***	1.164***	1.099***
	(0.023)	(0.021)	(0.023)	(0.024)
3rd quartile	1.793***	1.360***	1.406***	1.309***
	(0.031)	(0.026)	(0.026)	(0.026)
Top quartile	1.972***	1.498***	1.910***	1.727***
	(0.050)	(0.042)	(0.036)	(0.035)
Remedial courses taken at entry (ref. = none taken)				
Remedial math only	1.213***	0.869***	1.187*	1.615***
	(0.016)	(0.016)	(0.104)	(0.171)
Remedial English only	1.293***	0.882***	1.308	2.325***
	(0.020)	(0.019)	(0.198)	(0.450)
2 or more remedial classes	1.231***	0.880***	1.380	2.486**
	(0.019)	(0.019)	(0.420)	(0.842)
GPA semester 1 (ref. = D-/F: <1.00)				
D		1.086		1.617***
		(0.054)		(0.091)
D+		1.399***		1.902***
		(0.050)		(0.093)
C-		1.743***		2.509***

		(0.054)		(0.116)
C		1.989***		3.306***
		(0.071)		(0.153)
C+		2.421***		4.390***
		(0.076)		(0.201)
B-		3.025***		5.650***
		(0.089)		(0.258)
B		3.864***		7.542***
		(0.126)		(0.349)
B+		4.669***		8.751***
		(0.145)		(0.410)
A+/A		5.503***		10.421***
		(0.157)		(0.498)
No GPA record		2.479***		3.802***
		(0.097)		(0.274)
Credits earned semester 1 (ref. = 4 to 7 credits)				
0 credits		0.357***		0.866
		(0.014)		(0.064)
1 to 3 credits		0.515***		0.541***
		(0.009)		(0.033)
8 to 11 credits		1.659***		1.795***
		(0.034)		(0.058)
12 to 13 credits		2.536***		2.973***
		(0.070)		(0.097)
14 to 15 credits		3.249***		4.110***
		(0.118)		(0.140)
16 to 17 credits		4.279***		4.909***
		(0.224)		(0.183)
18 to 19 credits		4.262***		6.167***
		(0.484)		(0.471)
20 or more credits		2.955***		2.486***
		(0.841)		(0.544)
Remedial semester 1 (ref. = none taken)				
Failed/Withdrawn		0.751***		0.480***
		(0.019)		(0.076)
Passed		2.093**		0.890
		(0.037)		(0.091)
Constant	0.173***	0.155***	0.141***	0.057***
	(0.011)	(0.012)	(0.010)	(0.005)
Observations	263,857	263,857	277,915	277,915

Standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05

Appendix B4: Predicting graduation with variables at entry and first semester (Illinois)

Variables	AA at entry	
	Odds Ratios	
	Entry variables	Entry + semester 1 variables
ACT score (ref. = 17-24)		
ACT score 1-8	0.709 (0.337)	1.208 (0.623)
ACT score 9-16	0.764*** (0.028)	0.862*** (0.038)
ACT score 25-30	0.792 (0.115)	0.711* (0.120)
ACT score 31-36	0.458*** (0.022)	0.610*** (0.033)
Unknown	0.534*** (0.020)	0.594*** (0.026)
Workload semester 1 (ref. 12 to 13 credits)		
Less than 8 credits	0.206*** (0.031)	1.015 (0.167)
8 to 11 credits	0.397*** (0.028)	0.908 (0.071)
14 to 15 credits	1.524*** (0.027)	1.084** (0.033)
16 to 17 credits	2.197*** (0.051)	1.133*** (0.044)
18 to 19 credits	2.436*** (0.100)	0.935 (0.067)
20 or more credits	2.689*** (0.161)	0.919 (0.107)
Major selection at entry (ref. = liberal arts)		
General studies – AA		0.746*** (0.021)
General studies – AS		1.042 (0.025)
Others		0.860*** (0.019)
HS GPA percentile (ref. = 0 percentile rank)		
25% or less	1.069 (0.046)	1.150** (0.058)
26-50%	1.298*** (0.052)	1.172*** (0.053)

51-75%	1.375*** (0.056)	1.154** (0.053)
76-100%	1.271*** (0.056)	1.086 (0.056)
Unknown	1.068 (0.037)	1.052 (0.041)
Age at entry (ref. = 19)		
18 or younger	1.488*** (0.037)	1.293*** (0.037)
20	0.858*** (0.040)	0.895* (0.047)
21	0.750*** (0.046)	0.710*** (0.048)
22	0.860* (0.059)	0.695*** (0.053)
23	1.029 (0.072)	0.749*** (0.059)
24	0.914 (0.072)	0.626*** (0.055)
25 or older	1.164*** (0.043)	0.723*** (0.030)
Economic disadvantage status (ref. = no)		
Yes	0.823*** (0.014)	0.851*** (0.016)
Remedial requirement at entry (ref. = no remedial requirement)		
Remedial math required only	0.675*** (0.012)	
Remedial reading required only	0.529*** (0.029)	
Remedial writing required only	0.599*** (0.026)	
2 or more remedial requirements	0.379*** (0.009)	
GPA semester 1 (ref. = D-/F: <1.00)		
D: 1.00-1.32		1.325*** (0.107)
D+: 1.33-1.66		1.717*** (0.124)
C-: 1.67-1.99		3.489*** (0.214)
C: 2.00-2.32		2.602*** (0.176)
C+: 2.33-2.66		5.622***

		(0.342)
B-: 2.67-2.99		10.300***
		(0.620)
B: 3.00-3.32		8.629***
		(0.528)
B+: 3.33 - 3.66		15.318***
		(0.957)
A+/A: 3.67-4.00		18.125***
		(1.140)
Credits earned semester 1 (ref. = 12 to 13 credits)		
0 credits		0.338***
		(0.025)
1 to 3 credits		0.246***
		(0.018)
4 to 7 credits		0.390***
		(0.017)
8 to 11 credits		0.603***
		(0.017)
14 to 15 credits		1.294***
		(0.046)
16 to 17 credits		1.767***
		(0.083)
18 to 19 credits		2.101***
		(0.190)
20 or more credits		1.875***
		(0.264)
Remedial math semester 1 (ref. = not required, not taken)		
Required, not taken		1.003
		(0.022)
Passed all		1.432***
		(0.040)
Failed/withdrew one or more		1.444***
		(0.080)
Remedial English semester 1 (ref. = not required, not taken)		
Required, not taken		1.005
		(0.078)
Passed all		1.415***
		(0.029)
Failed/withdrew one or more		0.884**
		(0.041)
Constant	0.736***	0.124***
	(0.038)	(0.010)
Observations	81,600	81,600

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$