

Early Indicators of Student Success: A Multi-state Analysis

Paul Attewell^{1*}, Christopher Maggio¹, Frederick Tucker¹, Jay Brooks²,
Matt S. Giani³, Xiaodan Hu⁴, Tod Massa⁵, Feng Raoking⁵,
David Walling³, & Nathan Wilson²

¹ Sociology, CUNY Graduate Center, New York City

² Research and Policy Studies, Illinois Community College Board, Springfield

³ University of Texas at Austin, Austin

⁴ Northern Illinois University, DeKalb

⁵ State Council of Higher Education for Virginia, Richmond

Abstract

This paper reports the results of a four-state collaboration—Illinois, New York, Texas, and Virginia—that uses Student Unit Record Database Systems that track students from high school into college. The goal is to determine whether it is possible to accurately predict whether individual students will not graduate using very early indicators available at college entry or during the first semester. Using similar statistical models across four state university systems, we identify individual students at greatest risk of non-completion quite accurately at early stages, allowing college staff to prioritize interventions and supports aimed at improving completion for those at greatest risk. Our logistic regression models rely on variables available to university administrators at student entry, including high school GPA, standardized test scores, parental income, remediation requirements, declared major, and college credits attempted in the first semester. Our models do not use gender, race, or ethnicity in determining probability

* Contact: pattewell@gc.cuny.edu



of non-completion, making them useful for public university administrators. The fact that the same factors accurately predict graduation and non-completion in four very different state contexts suggests that similar dynamics are at play across the country. Our findings suggest that current commercial products that require extensive effort from faculty to input data on student progress, to act as an early warning system, may be unnecessary. More easily obtainable data can accurately predict students at risk of non-completion.

Keywords: degree performance, early indicators, prediction, student unit record database systems, undergraduates

Early Indicators of Student Success: A Multi-state Analysis

In recent years, several states have constructed very large databases that track multiple cohorts of students from high school into college and later into the labor force. These Student Unit Record Database Systems (SURDS)—also known as Student Unit Record Systems (SURS) or State Longitudinal Data Systems (SLDS)—typically compile socio-demographic information for hundreds of thousands of high school students in a state, along with academic test scores and high school GPA (Hearn et al., 2008). For those students who subsequently attend college, state-level SURDS compile information about colleges attended and transcript details of grades, credits, major field of study, and degrees (New America Foundation, 2017). Some database systems also track students' earnings during and after college, incorporating state unemployment insurance records. These databases are purged of individual identifiers and access is restricted to researchers under highly controlled circumstances to ensure data security.

SURDS compile their data from administrative sources, which are believed to be more accurate than self-reported data from surveys. The large number of cases in a SURDS enables researchers to focus on subpopulations of students, whether defined in terms of student characteristics or defined institutionally (e.g., those attending flagship versus secondary campuses of public universities). SURDS that document individuals over a long period allow scholars to observe different trajectories from high school through college and into the labor market and identify points in the educational process where students tend to fall behind, or to evaluate the labor market consequences of taking different routes through college.

With SURDS, higher education research has entered the era of Big Data. Dynarski and Berends (2015) have argued that SURDS will be a game changer for educational research. However, at present, these databases are collected and maintained separately by individual states, because federal agencies are forbidden by current law from using their administrative records to track student progress nationwide. Efforts to reverse this legal prohibition in the future may open the way to a national SURDS (Kreighbaum, 2017).

Purpose of Present Study

In this paper, we present findings from a multi-state research project, a collaboration of research teams from Illinois, New York, Texas, and Virginia, each analyzing their own SURDS to predict degree completion of individuals. The central research question asks whether it is possible to use SURDS data about undergraduates at the beginning of their college careers to accurately predict which individuals later fail to graduate. A second goal is to determine whether the same predictive model regarding graduation works equally well across these diverse states. To the extent that SURDS can accurately identify within the first or second semester of college which specific undergraduates

are at high risk of non-completion, college administrators may use this information to direct or prioritize counseling, support, or other interventions to improve undergraduates' chances of retention and graduation.

SURDS versus Early Alert Systems

Several commercial information systems (e.g., Starfish, EAB, Civitas), and some systems developed by colleges themselves—such as Purdue University's Course Signals—are currently used to provide warnings or early alerts that a particular undergraduate student is at risk of dropping out or failing a specific course (Massing et al., 2022). Early alert systems are college-based in contrast to state-level SURDS and typically require faculty members to input information about their students' class attendance and/or their grades on midterms and other assignments. Using prediction algorithms and these data from teachers, early alert software flags students at high risk of failure in a particular course, in principle allowing college staff to intervene with students in academic difficulty.

Much of the research on early alert systems focuses on the success of algorithms in accurately predicting student failure in a course (Liz-Dominguez et al., 2019; Massing et al., 2022). On this criterion, many of these systems perform well. Whether these alerts change student behavior and improve their academic success is another question. On this issue, the limited evidence is more mixed (Straumsheim, 2013). Massing and colleagues (2022), for example, studied whether sending warning emails to students resulted in changed success in a class, concluding: "Our results . . . do not provide any evidence that the warning mail has a significant effect on the results (or behavior) of the students" (p. 8).

In contrast to early alert systems, SURDS do not require special data collection efforts by faculty, instead using data already collected for each student at entry to college, along with institutional records on courses, course-loads, and grades. Where early alert systems focus on predicting passage or failure in a particular course or courses, our SURDS analyses predict degree completion. One practical rationale for the current study is therefore to assess whether the analysis of SURDS data is a viable and less resource-demanding alternative to commercial early alert systems for flagging students at high risk of academic failure and for allowing colleges to prioritize interventions and support services.

The analyses presented below present statistical models that estimate individual undergraduates' probability of graduating, using early SURDS indicators. Using logistic regression models, we find that these models yield accurate predictions of non-completion across four diverse states. The predictions are most accurate for those students at the highest risk of non-completion. We detail the contents of these models below and discuss their conceptual and methodological foundations. We also address

some important ethical concerns and practical challenges faced by colleges that wish to use these early indicators of student success to prioritize interventions and support services.

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, 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, constitutes a major hurdle to degree completion has led to widespread efforts to reform that part of the curriculum (Bailey et al., 2010; Chen & Simone, 2016; Hayward & Willett, 2014; Logue et al., 2019; Mokher & Hu, 2022). Other researchers, however, have emphasized that half of college drop-outs are in good academic standing when they leave, implying that the undergraduate retention problem is much wider than academic under-preparation (Abele, 2021).

Other research on retention emphasizes competing demands faced by undergraduates who need to juggle academic studies with paid employment and family obligations, creating time binds that lower graduation rates (Bozick, 2007; Stinebrickner & Stinebrickner, 2003, 2004). In addition, St. John (2003), Goldrick-Rab (2016), and others 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. Finally, an increasingly prominent theme in the student success literature focuses on institutional practices that contribute to low college completion, from financial aid policies (Baum & Scott-Clayton, 2013) and loss of credits after transferring (Monaghan & Attewell, 2015), to limited course availability and scheduling (Abele, 2021).

The literature on student success has identified many factors associated with attrition and completion and has estimated the average effects of those predictors across representative samples of students. The emphasis has not, however, been on accurately predicting completion outcomes for individual students to inform interventions, the goal of the present paper. For academic knowledge to translate into effective educational policy, it must move beyond the study of general trends. Seidman (2005) asserts that improvements to college retention rates require early identification of students at

risk of dropping out, coupled with early interventions aimed at aiding at-risk students, suggesting utilizing a “thorough examination of academic records’ available to colleges upon student entry, including grades, courses taken, and standardized test scores” (p. 21). Tucker and McKnight (2017) found success identifying students at risk for non-retention at a public university in the Midwest by examining high-school grades and ACT scores upon entry, findings bolstered further when examining GPA after one semester in college. Other models for identifying at-risk students rely on surveys that measure student academic, social, and psychological attachment to college life, including Baker and Siryk’s Student Adaptation to College Questionnaire (Baker & Siryk, 1986) and the recent Inventory of New College Student Adjustment developed by Watson and Lenz (2018). These surveys, however, do not rely on data already available to college administrators upon student entry.

Method

Data

This project linked researchers from four states who were interested in participating in a joint project, and who had access to and experience with their state’s postsecondary data systems (SURDS). Each state team undertook separate data analyses based on their own state’s SURDS. For security and legal reasons, no data was shared or moved across states; instead, each state team followed a common protocol for statistical analyses using their own state’s data. Only statistical outputs were shared. The selection of the four states—Illinois, New York, Texas, and Virginia—was not random or representative in any way. Each had a well-established SURDS and researchers willing to participate. This unusual collaboration was funded by the Bill & Melinda Gates Foundation, in part to establish whether this kind of multi-state collaborative research was practical and useful.

The researchers analyzed undergraduate cohorts that entered college from Fall 1999 through Spring 2010 and were followed through 2016. 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 SURDS, private colleges did not usually report transcript information, so we excluded them from the analyses reported in this paper. The SURDS were longitudinal, following each student from entry into college for at least ten years, at which time indicating whether an individual had graduated. Sample sizes ranged from 48,783 Bachelor of Arts (BA) students in Texas, to 357,836 for Associate in Arts (AA) students in New York.

Variables

Graduation was constructed as a binary outcome variable. For students initially entering a four-year college, we defined our milestone as completion of the bachelor’s

degree within 12 semesters of entry. The equivalent milestone for undergraduates who started at community colleges was more complicated. Students who begin at community college often say they intend to earn their baccalaureate, and many transfer to a four-year college without first completing their associate degree (Long & Kurlaender, 2009). Thus, simply counting whether a student received an associate degree is misleading as a measure of community college student success. Instead, we counted as having reached an important milestone those community college matriculants who either obtained an associate or baccalaureate degree or had accumulated 60 or more credits, which is the minimum required at most community colleges to receive an associate degree.

After exploratory modeling, we settled on the following independent variables available at the beginning of each student's freshman year: age at college entry; parental adjusted gross income (AGI); high school GPA, SAT, ACT, or Texas Assessment of Knowledge and Skills (TAKS) score; remedial requirements (math, reading, writing); workload in the first semester (total number of credits, counting both remedial and non-remedial courses); and whether a major was declared at college entry (a dichotomous variable). Another set of variables contained measures of student performance during the first semester of college: GPA in first semester; credits earned in first semester; whether remedial math was taken in the first semester (if required), and if so whether it was passed; and equivalent measures for taking or passing remedial reading or remedial writing in the first semester. All continuous variables were converted into categorical predictors, allowing for the addition of a 'missing' category for each variable.

Table 1 reports details for each variable. 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 its high school seniors, transforming this into a percentile score, whereas the New York data used SAT scores. Different state SURDS 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, although they are quite similar. We also examined whether adding measures that described student performance after the first year of college improved the accuracy of models predicting graduation. We found that later academic performance measures did not substantially improve prediction. Combined with SURDS items, data on how well a student performed during their first semester of college were sufficiently predictive, as documented below. The specific logistic regression equations are available in an online appendix using the following link: <https://inequalityinhighered.org/2019/12/03/early-indicators-of-student-success-a-multi-state-analysis/>.

Table 1. Variables Description (New York)

Dependent variable	
Graduation (AA, BA, or 60 credits for AA entrants)	
Independent variables	
Age at entry	High school GPA
18 or younger	A-/A: 3.67–4.00
19	B+: 3.33–3.67
20	B: 3.00–3.33
21	B-: 2.67–3.00
22	C+: 2.33–2.67
23	Parental adjusted gross income
24	1 st quartile (highest)
25 or older	2 nd quartile
SAT score	3 rd quartile
1 st quintile	4 th quartile
2 nd quintile	Parent AGI missing
3 rd quintile	Parent AGI missing for cohort
4 th quintile	Remedial requirements at entry
5 th quintile	No remedial requirement
No SAT score	Remedial math required only
Major in semester 1	Remedial reading required only
Declared	Remedial writing required only
Not declared	Two or more remedial requirements
Unclassified (unknown)	Remedial requirement unknown
Workload semester 1	Remedial math semester 1
< 8 credits	Not required, not taken
≥ 8 credits & < 12 credits	Required, not taken
≥ 12 credits & < 14 credits	Passed all
≥ 14 credits & < 16 credits	Failed/withdrew one or more
≥ 16 credits & < 18 credits	Remedial reading semester 1
≥ 18 credits & < 20 credits	Not required, not taken
> 20 credits	Required, not taken
Remedial writing semester 1	Passed all
Not required, not taken	Failed/withdrew one or more
Required, not taken	GPA semester 1 (non-remedial)
Passed all	A-/A: 3.67–4.00
Failed/withdrew one or more	B+: 3.33–3.67
Credits earned semester 1	B: 3.00–3.33
0 credits (but enrolled)	B-: 2.67–3.00
> 0 credits & < 4 credits	C+: 2.33–2.67
≥ 4 credits & < 8 credits	C: 2.00–2.33
≥ 8 credits & < 12 credits	C-: 1.67–2.00
≥ 12 credits & < 14 credits	D+: 1.33–1.67
≥ 14 credits & < 16 credits	D: 1.00–1.33
≥ 16 credits & < 18 credits	D-/F: < 1.00
≥ 18 credits & < 20 credits	Enrolled, no GPA record
20 credits or more	

Avoiding Algorithmic Bias

Readers will note that we deliberately eschew using gender, race, or ethnicity as predictors of graduation in the statistical models that follow. Despite being aware that these demographic characteristics are on average associated with higher or lower completion, we avoided building models predicting graduation that relied on group characteristics of this type. To do so might reify stereotypes and lead to what economists term ‘statistical discrimination’—assessing individuals’ promise by their group membership, resulting in disadvantages for those individuals whose performance differs from their group average (Arrow, 1973). More recently, the potential for bias in computerized decision-making has been conceptualized as ‘Algorithmic Bias’ (Baer, 2019; Baker & Hawn, 2021; Government Accountability Office, 2022; Noble, 2018; O’Neil, 2016). The core idea is that ostensibly fair or neutral computer programs used to make decisions may nevertheless contain features in their algorithms which can systematically disadvantage one group of persons compared to another. In computerized decisions about bail, for example, Black defendants have been shown to have a higher error rate of being incorrectly classified as likely to reoffend and therefore are more likely than Whites to be denied bail (Corbett-Davies et al., 2017). Baker and Hawn (2021) provide a review of the terminology, concepts, and literature regarding algorithmic bias in education.

Given these concerns regarding algorithmic bias, we constructed models that omit gender, race, and ethnicity as predictors of graduation. As a check, after completing our analyses without such attributes, we determined separately for each state’s SURDS 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. There were two demographic exceptions: a student’s age at college entry and a family income measure were both associated with graduation such that omitting those predictors would impair predictive power. We judged that incorporating those two variables into the models would be less problematic than building predictive models in which gender, race, or ethnicity played a substantial role.

Statistical Models

We followed a multi-stage strategy for creating and evaluating predictive models. Initially, we randomly assigned the undergraduate sample into two different parts: training and test samples (Rogers & Girolami, 2012). Our training data, consisting of 70% of the randomly selected cases, was used to construct a predictive logistic regression model. The remaining 30% of the full sample, called test data, was withheld from the logistic regression and used in a final step to evaluate the predictive accuracy of the regression model.

After a logistic regression model was estimated from the training data, the prediction equation obtained from that analysis was applied to score the cases in the unused test sample, providing a predicted probability (or \hat{p}) of reaching the graduation milestone for each individual in the test group. We standardized the distribution of \hat{p} -hats into decile groups of equal size. To evaluate the model’s accuracy in its tails, we focused

on the bottom of this distribution: those cases for which our model indicated a very low chance of graduation. The predicted probabilities calculated from our model were then compared (cross-tabulated) with the actual measured milestone outcomes for the test-sample individuals, yielding validation statistics that measured the accuracy of our predictions for the test sample.

Cross-validation assesses whether a model developed from a training sample generalizes to out-of-sample test data, reflecting the overall population, and therefore can be considered reproducible. Cross-validation is a protection from what data miners term ‘overfitting’—the possibility that a strongly predictive model partly reflects random noise or finds relationships in a dataset that would not apply to data drawn from other samples (Rogers & Girolami, 2012). In all our tables presented below, the reported accuracy statistics are always for the held-back test data, which is the equivalent of applying the predictive model to new data.

Findings

Generally, the most powerful predictors of graduation in the regression analyses were consistent with the academic momentum perspective: the number of credits earned in the first semester and first semester GPA were the strongest predictors of long-term graduation. 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 additional important factors associated with graduation in baccalaureate programs.

Table 2 reports these predicted probability statistics for each state’s two-year entrants, with Model A including only variables known at entry, and Model B adding first-semester variables. The resulting pattern is in every case strongly curvilinear: the predictive accuracy of the model is very high for those students least likely to graduate. For students in the highest 1% group of risk scores at entry to community college, non-graduation rates ranged from 85% for Virginia, to 94% for Texas in Model A, and above 97% for all four states in Model B.

Even for students with the highest 20% of risk scores, non-graduation accuracy ranged from 79% for New York, to 88% for Virginia in Model A, and above 91% for all four states in Model B. At the other end of the spectrum, those students with the lowest decile of risk scores—those most likely to graduate—did not graduate at a range of 15% in Texas, to 44% in Virginia in Model A. For students with predicted graduation probabilities in the mid-range, the accuracy of these regression models is much lower. Students in most middle deciles, for example, have close to average graduation rates.

Table 2. 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 Model A = Variables at college entry. ^b Model B = Variables at college entry + first semester variables.

A similar pattern emerges among four-year college entrants (Table 3), albeit with higher graduation rates typical for these students. For students in the highest 1% group of risk scores at entry to a four-year college, the non-graduation rate ranged from 76% for Virginia, 87% for New York, and 92% for Texas in Model A.

With the addition of first-semester variables (Model B in Table 3), the non-graduation accuracy for the highest-risk baccalaureate students rose to 95% and above. For students in the lowest decile of risk, only 7% did not graduate in Virginia, 14% did not graduate in Texas, and 21% did not graduate in New York for Model A. These predictions were more accurate after the first semester's data was added in Model B, with non-graduation rates ranging from 4.5% for Virginia's lowest decile of risk, to 16% for New York. Like with the two-year entrants, the students with middle-decile risk scores were more of a 50:50 proposition.

If our goal were accurate prediction for every student across the distribution, this curvilinear distribution would be a serious drawback. When the goal, however, is to provide

Table 3. 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 (<i>N</i> =47,045)		Texas (<i>N</i> =20,736)		Virginia (<i>N</i> =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 Model A = Variables at college entry. ^b Model B = Variables at college entry + first semester variables.

actionable information identifying those students at highest risk of not graduating, the logistic models successfully identify those students of highest risk of non-graduation. This leads us to consideration of how a college might use such scores and considers ethical issues such as the potential harm resulting from mistakenly identifying individuals as high risk (i.e., false positives).

Discussion

The predictive accuracy of our statistical models has a curvilinear shape for all four state SURDS. The highest predictive accuracy—usually exceeding 95 percent—occurs for students at the highest risk of non-graduation with models that incorporate first semester variables. Prediction in the middle of the distribution is much less accurate. In

this ‘murky middle’ exists a large swath of students who have roughly similar chances of graduating or not graduating. These early indicator models are *not* effective in distinguishing among those in the middle.

The Logic of Prioritizing Intervention and Support Services

The existence of a murky middle does not lessen the value of the early indicator models if the purpose of intervening is to enhance graduation rates. Knowing which among their incoming students are most and least likely to graduate, provided in a timely fashion by our early indicator models in the form of risk scores, would allow college staff to prioritize outreach to students and target support services to those at greatest need. In our view, prioritization of interventions and support services is the most immediate and practical use of our early indicator models of student success.

It is possible that some institutions might use predictive scores to separate matriculants at highest risk of non-completion from other students at less risk, and tailor a special program for the former. One analogous situation is the City University of New York’s pre-matriculation program, CUNY Start. This voluntary program identifies applicants to community college who have multiple remediation needs according to placement-test scores in math, reading, and writing, taken just before they intend to start college (City University of New York [CUNY], 2017), only a fraction of the data we use in our multivariate predictive models. Identified students at CUNY are invited to defer immediate enrollment in a community college program, and instead are offered the option of taking one or two twelve-week courses that focus on remedial coursework, at very low tuition, taught by teachers especially skilled at adult education (CUNY, 2017). The goal of this program is to raise students’ skills to such a level that they can pass the skills tests and begin their community college program without further remedial coursework.

Initial evaluations of the CUNY Start program report that significantly larger proportions of Start students pass the skills tests needed to exit remediation than a comparison group of community college students who take remedial coursework alongside non-remedial classes, during their early semesters at community college (Scrivener & Logue, 2016). Some students who do not pass their courses in this special track may decide not to enter community college—we do not have data on how many—and if so, would have paid far less tuition than they would have done had they started community college and taken remedial coursework there. One should recall that Adelman (2006) found that about 13% of entering students in a national sample drop out before completing ten credits.

A second policy option would use risk scores to identify high-risk students in order to offer those students additional academic or social supports (cf. Tinto, 2012). Another analogue can be found in the City University of New York’s Accelerated Study in Associate Programs (ASAP), which identifies potential participants based on only a fraction of the data we utilize: low placement-test scores that cause students to take one or two

remedial courses. ASAP offers selected participants a range of extra supports, such as individual advisors or counsellors whom they meet on a regular basis, block schedules, tracking into courses with other similarly situated students, and material benefits like free textbooks and transportation. CUNY's ASAP program is a complex intervention, initially targeted at high-risk community college students. Random assignment evaluations have documented near-doubling of graduation rates for ASAP students compared to control groups (Gupta, 2017).

In 2015, Ohio began replicating this ASAP program in three of its community colleges. Student volunteers were randomly assigned treatment of academic services similar to CUNY's, as well as financial assistance in the form of tuition and textbook waivers, and career advisement. Ohio's ASAP program doubled the graduation rate for students with developmental requirements and significantly boosted graduation rates for those without. This program cost the colleges an additional 42% per student but cost 22% less per degree conferred compared with the control group (Miller et al., 2020). These two programs illustrate two types of policies where participants are selected using early or leading indicators of student success. We stress, however, that these two real-world examples did not use the multivariate indicators discussed earlier in this paper, instead selecting participants on a voluntary basis, based on the placement test scores of incoming undergraduates.

This brings us to the important issue of how error in prediction might affect such policies regarding interventions. The first thing to note is that the current indicators of risk being used by many colleges, specifically skills or placement test scores from commercially available ACCUPLACER or COMPASS tests, have been criticized for being inaccurate (Rodriguez et al., 2014; Scott-Clayton, 2012). The status quo approach to identifying at-risk students and directing them into special classes is already error laden.

For the higher-risk end of the spectrum, most predictions of non-graduation in the models we developed were 95% accurate or better, meaning that at most five in a hundred identified as being at risk of non-completion would in fact have completed their degree. One ethical issue is whether individuals (or colleges) would be harmed if these 5% were erroneously classified as high-risk. If risk scores are used to prioritize provision of extra academic and counselling support, it seems unlikely that a misclassified student would be harmed by being encouraged to make use of such targeted support, especially since students may decide to spurn those supports if they so choose. From the institution's perspective, if they identify students for extra support, 95% of whom would not be likely to complete their degree and (due to inaccurate prediction) 5% would graduate anyway, then perhaps 5% of the extra supports are wasted in the sense that they would better be targeted elsewhere. In our judgement, this is a relatively small misallocation of resources, with little risk of harm to students who were misclassified as needing those supports. More serious harm would occur if risk scores were used to discourage students deemed at-risk from attempting a degree program, since about 5% of those identified persons would have completed a degree. We would argue against that type of use for early indicators.

There are additional rationales for using early indicators to target institutional interventions and support. In the past, counseling, tutoring, and other academic support services have often followed a ‘first-come, first-serve’ approach, leaving it to each student facing difficulties to come and ask for help. Unfortunately, many students who need help do not seek it. Researchers have shown, for example, that young Black men tend not to make full use of academic support systems (Bush & Bush, 2010). Both early alert systems and the SURDS approach suggested by this paper take a proactive or intrusive approach to academic support: the institution reaches out to specific students who seem to be facing difficulties. That model focuses resources on those most in need and prioritizes outreach and intrusive advisement over waiting for students to seek help.

This research collaboration has shown that it is possible to develop early indicators of student success and non-completion and to link those early indicators to interventions. As more states provide access to SURDS data, we expect to see wider use of early indicators of student success as a way of targeting support services and for assessing institutional interventions aimed at improving student success.

Author Note

We have no known conflicts of interest to disclose. This work was funded by grants from the Bill & Melinda Gates Foundation (Grant # OPP1159855) and Ascendium Education Solutions, formerly the Great Lakes Higher Education Guaranty Corporation (Grant # G-201704–15499).

References

- Abele, L. (2021). Institutional barriers contribute to low college completion rates. *Journal of Postsecondary Student Success*, 1(1), 18–24. https://doi.org/10.33009/fsop_jpss124555
- Adelman, C. (1999). *Answers in the toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment*. <https://eric.ed.gov/?id=ED431363>
- Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. U.S. Department of Education. <https://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/toolbox.pdf>
- Arrow, K. J. (1973). The theory of discrimination. In O. Aschenfelter & A. Rees (Eds.), *Discrimination in labor markets* (pp. 3–33). Princeton University Press.
- Baer, T. (2019). *Understand, manage, and prevent algorithmic bias: A guide for business users and data scientists*. Apress.

- Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, 29(2), 255–270. <https://doi.org/10.1016/j.econedurev.2009.09.002>
- Baker, R. S., & Hawn, A. (2021, March 1). *Algorithmic bias in education*. <https://doi.org/10.35542/osf.io/pbmvz>
- Baker, R., & Siryk, B. (1986). Exploratory intervention with a scale measuring adjustment to college. *Journal of Counseling Psychology*, 33(1), 31–38. <https://doi.org/10.1037/0022-0167.33.1.31>
- Baum, S., & Scott-Clayton, J. (2013). Redesigning the Pell Grant program for the twenty-first century. *The Hamilton Project, Discussion Paper 2013–04*. Brookings Institute. <https://vtechworks.lib.vt.edu/bitstream/handle/10919/90858/RedesigningPellGrantProgram.pdf?sequence=1>
- Bozick, R. (2007). Making it through the first year of college: The role of students' economic resources, employment, and living arrangements. *Sociology of Education*, 80(3), 261–285. <https://doi.org/10.1177%2F003804070708000304>
- Bush, E. C., & Bush, L. (2010). Calling out the elephant: An examination of African American male achievement in community college. *Journal of African American Males in Education*, 1(1), 40–62.
- Chen, X., & Simone, S. (2016). *Remedial course taking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes*. National Center for Education Statistics (NCES 2016-405). U.S. Department of Education. <https://nces.ed.gov/pubs2016/2016405.pdf>
- Chingos, M. (2018). *What matters most for college completion? Academic preparation is the key predictor of success*. American Enterprise Institute.
- City University of New York. (2017). CUNY Start. <http://www2.cuny.edu/academics/academic-programs/model-programs/cuny-college-transition-programs/cuny-start/>
- Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 797–806). <https://doi.org/10.1145/3097983.3098095>
- Dynarski, S., & Berends, M. (2015). Introduction to the special issue. *Education Evaluation and Policy Analysis*, 37(1), 3s–5s. <https://doi.org/10.3102/0162373715575722>
- Goldrick-Rab, S. (2016). *Paying the price: College costs, financial aid, and the betrayal of the American dream*. University of Chicago Press.
- Government Accountability Office. (2022). *Consumer protection: Congress should consider enhancing protections around scores used to rank consumers*

- (GAO-22-104527). United States Government Accountability Office. <https://www.gao.gov/assets/gao-22-104527.pdf>
- Gupta, H. (2017). *The power of fully supporting community college students: The effects of the City University of New York's Accelerated Study in Associate Programs after six years*. MDRC. <https://www.mdrc.org/publication/power-fully-supporting-community-college-students>
- Hayward, C., & Willett, T. (2014). *Curricular redesign and gatekeeper completion: A multicollege evaluation of the California Acceleration Project*. The RP Group. <http://cap.3csn.org/files/2014/04/RP-Evaluation-CAP.pdf>
- Hearn, J., McLendon, M., & Mokher, C. (2008). Accounting for student success: An empirical analysis of the origins and spread of state student unit-record systems. *Research in Higher Education*, 49(8), 665–683. <https://doi.org/10.1007/s11162-008-9101-z>
- Kreighbaum, A. (2017, May 16). Push for unit records revived. *Inside Higher Ed*. <https://www.insidehighered.com/news/2017/05/16/bipartisan-bill-would-overturn-federal-ban-student-unit-record-database>
- Liz-Dominguez, M., Caeiro Rodriguez, M., Llamas Nistal, M., & Mikic Fonte, F. (2019). Predictors and early warning systems in higher education: A systematic literature review. *Applied Sciences*, 9(24), 5569. <https://doi.org/10.3390/app9245569>
- Logue, A. W., Douglas, D., & Watanabe-Rose, M. (2019). Corequisite mathematics remediation: Results over time and in different contexts. *Educational Evaluation and Policy Analysis*, 41(3), 294–315. <https://doi.org/10.3102/0162373719848777>
- Long, B. T., & Kurlaender, M. (2009). Do community colleges provide a viable pathway to a baccalaureate degree? *Educational Evaluation and Policy Analysis*, 31(1), 30–53. <https://doi.org/10.3102/0162373708327756>
- Massing, T., Reckmann, N., Klenke, J., Otto, B., Hanck, C., & Geodicke, M. (2022). Effects of early warning emails on student performance. *arXiv*. <https://doi.org/10.48550/arXiv.2102.08803>
- Miller, C., Headlam, C., Manno, M., & Cullinan, D. (2020). *Increasing community college graduation rates with a proven model: Three-year results from the Accelerated Study in Associate Programs (ASAP) Ohio demonstration*. MDRC. https://www.mdrc.org/sites/default/files/ASAP_OH_3yr_Impact_Report_1.pdf
- Mokher, C., & Hu, S. (2022). Diverging paths: Exploring the association between initial math pathways and college students' subsequent math performance. *Journal of Postsecondary Student Success*, 1(3), 50–74. https://doi.org/10.33009/fsop_jpss129846
- Monaghan, D., & Attewell, P. (2015). The community college route to the BA. *Education Evaluation and Policy Analysis*, 37(1), 70–91. <https://doi.org/10.3102/0162373714521865>

- New America Foundation. (2017). *Student unit record data system*. <https://www.newamerica.org/education-policy/topics/higher-education-data-and-transparency/higher-education-data/student-unit-record-data-system/>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Books.
- Rodriguez, O., Bowden, B., Belfield, C., & Scott-Clayton, J. (2014). Remedial placement testing in community colleges: What resources are required, and what does it cost? *Community College Research Center (Working Paper No. 73)*. Teacher's College, Columbia University. <https://ccrc.tc.columbia.edu/media/k2/attachments/remedial-placement-testing-resources.pdf>
- Rogers, S., & Girolami, M. (2012). *A first course in machine learning*. CRC Press.
- Scott-Clayton, J. (2012). Do high-stakes placement exams predict college success? *Community College Research Center (Working Paper No. 41)*. Teacher's College, Columbia University. ccrc.tc.columbia.edu/media/k2/attachments/high-stakes-predict-success.pdf
- Scrivener, S., & Logue, A. (2016). *Building college readiness before matriculation: A preview of a CUNY Start evaluation*. MDRC. https://www.mdrc.org/sites/default/files/Building_College_Readiness_2016.pdf
- Seidman, A. (2005). Minority student retention: Resources for practitioners. *New Directions for Institutional Research*, 125, 7–24. <https://doi.org/10.1002/ir.136>
- Stinebrickner, R., & Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of Labor Economics*, 21(2), 473–491. <https://doi.org/10.1086/345564>
- Stinebrickner, R., & Stinebrickner, T. R. (2004). Time-use and college outcomes. *Journal of Econometrics*, 121(1/2), 243–269. <https://doi.org/10.1016/j.jeconom.2003.10.013>
- St. John, E. P. (2003). *Refinancing the college dream: Access, equal opportunity, and justice for taxpayers*. Johns Hopkins University Press.
- Straumsheim, C. (2013, November 6). Mixed signals. *Inside Higher Ed*. <https://www.insidehighered.com/news/2013/11/06/researchers-cast-doubt-about-early-warning-systems-effect-retention>
- Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *The Journal of Higher Education*, 59(4), 438–455. <https://doi.org/10.1080/00221546.1988.11780199>
- Tinto, V. (1994). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press.
- Tinto, V. (2012). *Completing college: Rethinking institutional action*. University of Chicago Press.

- Tucker, L., & McKnight, O. (2017). Assessing the validity of college success indicators for the at-risk student: Toward developing a best-practice model. *Journal of College Student Retention: Research, Theory & Practice*, 21(2), 166–183. <https://doi.org/10.1177/1521025117696822>
- Watson, J. C., & Lenz, S. (2018). Development and evaluation of the inventory of new college student adjustment. *Journal of College Student Retention: Research, Theory & Practice*, 22(3), 425–440. <https://doi.org/10.1177/1521025118759755>