AGENDA

FACULTY SENATE
Wednesday, November 20, 2019, 3 p.m.
Altgeld Hall 315
DeKalb, Illinois

I. CALL TO ORDER

II. ADOPTION OF THE AGENDA

III. APPROVAL OF THE OCTOBER 30, 2019 MINUTES – Pages 3-5

IV. PRESIDENT'S ANNOUNCEMENTS

V. ITEMS FOR FACULTY SENATE CONSIDERATION

A. Graduate School GRE Admissions Exams
Brad Bond, Dean, The Graduate School

VI. CONSENT AGENDA

VII. UNFINISHED BUSINESS

A. Proposal to amend Faculty Senate Bylaws, Article 3.2
Faculty Rights and Responsibilities Committee – Page 6
SECOND READING/ACTION

B. Prioritizing faculty issues – Pages 7-11

C. Proposed Admissions Policy update – Pages 12-202

VIII. NEW BUSINESS

IX. REPORTS FROM ADVISORY COMMITTEES

A. Faculty Advisory Council to IBHE – Linda Saborío – report

B. University Advisory Committee to the Board of Trustees – report
Jeffry Royce, Cathy Doederlein, Kendall Thu
Alex Gelman, Sarah Marsh, Jason Hanna

X. REPORTS FROM STANDING COMMITTEES

A. Faculty Rights and Responsibilities Committee – Katy Jaekel, Chair – no report

B. Academic Affairs Committee – Peter Chomentowski, Chair – no report

C. Rules, Governance and Elections Committee – Keith Millis, Liaison/Spokesperson – report

D. Resources, Space and Budget Committee – George Slotsve, Liaison/Spokesperson – report

XI. PUBLIC COMMENT

XII. INFORMATION ITEMS

A. Policy Library – Comment on Proposed Policies (right-hand column on web page)
B. Minutes, Academic Planning Council
C. Minutes, Athletic Board
D. Minutes, Baccalaureate Council
E. Minutes, Board of Trustees
F. Minutes, Campus Security and Environmental Quality Committee
G. Minutes, Comm. on the Improvement of the Undergraduate Academic Experience
H. Minutes, General Education Committee
I. Minutes, Graduate Council
J. Minutes, Graduate Council Curriculum Committee
K. Minutes, Honors Committee
L. Minutes, Operating Staff Council
M. Minutes, Student Senate
N. Minutes, Supportive Professional Staff Council
O. Minutes, University Assessment Panel
P. Minutes, University Benefits Committee
Q. Minutes, Univ. Comm. on Advanced and Nonteaching Educator License Programs
R. Minutes, University Committee on Initial Educator Licensure
S. 2019-20 Faculty Senate meeting dates:
Sep 4, Oct 2, Oct 30, Nov 20, Jan 22, Feb 19, Mar 25, Apr 22

XIII. ADJOURNMENT

MINUTES

FACULTY SENATE MEETING
Wednesday, October 30, 2019, 3 p.m.
Holmes Student Center Sky Room
Northern Illinois University
DeKalb, IL

Full transcript

VOTING MEMBERS PRESENT: Arado, Bateni, Bowers (for Shi), Buck, Burton, G. Chen, J. Chen, Chmaissem, Chomentowski, Collins, Demir, Duffin, Dzurian (for Riley), Farrell, Fredericks, Hanna, Hua, Jaekel, Jong, Keddie, Kim, Koss, Kuehl, Littauer, Macdonald, Mayer, McCarthy, McGowan, Millhorn, Montana, Nelson, Newman, Pendergrass, Polansky,k Powell, Qin, Reeves, Schatteman, Schraufnagel, Schuller, Sharp, Siegesmund, Skarbinski, Slotsve, Surjadi, Tatara, Thu, Un, Vahabzadeh, Villanueva, Weffer, Whedbee, Wilson, Zheng

VOTING MEMBERS ABSENT: Beyer, Bujarski, Creed, Grund, Hanley, Johnston-Rodriguez, Konen, Kot, Lampi, Martin, Millis, Mooney, Moraga, Penrod, Petgas, Rau, Riley, Scherer, Shi, Sirotkin, Staikidis, Subramony

OTHERS PRESENT: Bryan, Doederlein, Falkoff, Groza, Ingram, Jensen, Klaper, White, Whitelaw

OTHERS ABSENT: Ferguson, Gelman, Hanna, Marsh, Kortegast, Woodruff

I. CALL TO ORDER

Faculty Senate President **K. Thu** called the meeting to order at 3 p.m.

II. ADOPTION OF THE AGENDA

G. Slotsve moved to adopt the agenda, seconded by **P. Skarbinski.** Motion passed.

III. APPROVAL OF THE OCTOBER 2, 2019 MINUTES

R. Siegesmund moved to approve the minutes, seconded by **K. Jaekel**. Motion passed.

IV. PRESIDENT'S ANNOUNCEMENTS

A. Support for Undocumented Students at NIU

B. U.S. News & World Report ranking of NIU

C. National Association of Faculty Senates

V. ITEMS FOR FACULTY SENATE CONSIDERATION

A. Undergraduate Admissions Testing Optional
Beth Ingram, Executive Vice President and Provost
Sol Jensen, Vice President, Division of Enrollment Management, Marketing and Communications

G. Slotsve moved to ask the Baccalaureate Council to take up the topic of moving undergraduate admissions standards to less reliance on standardized tests and bring proposed implementation details back to Faculty Senate for review; seconded by **S. Weffer**. Motion passed.

41 – Yes
6 – No
2 – Abstain

B. Create an Accessible Syllabus
Katy Whitelaw, IT Accessibility Officer

C. Prioritizing faculty issues

VI. CONSENT AGENDA

VII. UNFINISHED BUSINESS

A. Proposal to amend Faculty Senate Bylaws, Article 3.4
Committee on the Economic Status of the Profession
SECOND READING/ACTION

O. Chmaissem moved to approve the proposal, seconded by **R. Siegesmund**. Motion passed.

44 – Yes
4 – No
1 - Abstain

VIII. NEW BUSINESS

A. Proposal to amend Faculty Senate Bylaws, Article 3.2
Faculty Rights and Responsibilities Committee
FIRST READING

IX. REPORTS FROM ADVISORY COMMITTEES

A. Faculty Advisory Council to IBHE – Linda Saborío – report

B. University Advisory Committee to the Board of Trustees – no report
Jeffry Royce, Cathy Doederlein, Kendall Thu
Alex Gelman, Sarah Marsh, Jason Hanna

X. REPORTS FROM STANDING COMMITTEES

A. Faculty Rights and Responsibilities Committee – Katy Jaekel, Chair – no report

B. Academic Affairs Committee – Peter Chomentowski, Chair – no report

C. Committee on the Economic Status of the Profession – no report

D. Rules, Governance and Elections Committee – Keith Millis, Liaison/Spokesperson – report

E. Resources, Space and Budget Committee – George Slotsve, Liaison/Spokesperson – report

XI. PUBLIC COMMENT

XII. INFORMATION ITEMS

A. Policy Library – Comment on Proposed Policies (right-hand column on web page)
B. Minutes, Academic Planning Council
C. Minutes, Athletic Board
D. Minutes, Baccalaureate Council
E. Minutes, Board of Trustees
F. Minutes, Campus Security and Environmental Quality Committee
G. Minutes, Comm. on the Improvement of the Undergraduate Academic Experience
H. Minutes, General Education Committee
I. Minutes, Graduate Council
J. Minutes, Graduate Council Curriculum Committee
K. Minutes, Honors Committee
L. Minutes, Operating Staff Council
M. Minutes, Student Senate
N. Minutes, Supportive Professional Staff Council
O. Minutes, University Assessment Panel
P. Minutes, University Benefits Committee
Q. Minutes, Univ. Comm. on Advanced and Nonteaching Educator License Programs
R. Minutes, University Committee on Initial Educator Licensure
S. 2019-20 Faculty Senate meeting dates:
Sep 4, Oct 2, Oct 30, Nov 20, Jan 22, Feb 19, Mar 25, Apr 2

XIII. ADJOURNMENT

It was moved and seconded to adjourn. Motion passed. Meeting adjourned at 4:26 p.m.

Proposed amendment to Faculty Senate Bylaws Article 3.2

1) The Faculty Senate Committee on the Economic Status of the Profession has been eliminated. 2) Economic issues for faculty that fall outside the Collective Bargaining Agreement can be brought to the Faculty Rights and Responsibilities Committee.

ARTICLE 3: STANDING COMMITTEES OF THE FACULTY SENATE

3.2 Faculty Rights and Responsibilities Committee

3.2.2 Duties

The committee shall advise the Senate on matters and issues concerning:

Representation of the faculty in the governance of the university;

Faculty compensation and benefits not covered by the Collective Bargaining Agreement;

Faculty participation in the development of university policies, procedures, and practices which advance the academic mission of the university and a learning environment throughout the university;

Collective and individual faculty prerogatives in university policies and procedures;

Standards and procedures of accountability concerning faculty ethics and responsibilities and adherence to those standards and responsibilities;

The climate of academic freedom for the university community and policies, procedures, and practices of the university as they affect academic freedom;

Specific academic freedom issues which warrant Senate attention;

The administration and effectiveness of the faculty grievance processes.

Rank up to five issues by checking the appropriate choice box. Only one issue per choice.

Issue	1st choice 5 points	2nd choice 4 points	3rd choice 3 points	4th choice 2 points	5th choice 1 point	TOTAL
Work and community, Annie Glidden North	4 votes 20 pts	1 vote 4 points	3 votes 9 points	2 votes 4 points	0 votes 0 points	37 points
Safety, anti-violence/sexual justice	2 votes 10 points	2 votes 8 points	5 votes 15 points	3 votes 6 points	4 votes 4 points	43 points
Deferred Action for Childhood Arrivals (DACA)	0 votes 0 points	1 vote 4 points	1 vote 3 points	1 vote 2 points	1 vote 1 point	10 points
Textbook costs	1 vote 5 points	1 vote 4 points	1 vote 3 points	5 votes 10 points	3 votes 3 points	25 points
Mental health	2 votes 10 points	2 votes 8 points	5 votes 15 points	2 votes 4 points	2 votes 2 points	39 points
Job classification	0 votes 0 points	1 vote 4 points	0 votes 0 points	1 vote 2 points	1 vote 1 point	7 points
De-centralizing budget	3 votes 15 points	5 votes 20 points	7 votes 21 points	9 votes 18 points	0 votes 0 points	74 points
Messaging/image of NIU	1 vote 5 points	4 votes 16 points	4 votes 12 points	3 votes 6 points	4 votes 4 points	43 points
Equity gap	7 votes 35 points	5 votes 20 points	4 votes 12 points	4 votes 8 points	3 votes 3 points	78 points
Space allocation	1 vote 5 points	2 votes 8 points	1 vote 3 points	2 votes 4 points	0 votes 0 points	20 points
Enrollment/recruitment and retention	18 votes 90 points	11 votes 44 points	6 votes 18 points	1 vote 2 points	6 votes 6 points	160 points
Food insecurity	0 votes 0 points	1 vote 4 points	0 votes 0 points	1 vote 2 points	5 votes 5 points	11 points
Reducing number of committees	1 vote 5 points	3 votes 12 points	2 votes 6 points	3 votes 6 points	8 votes 8 points	37 points
Enhancing diversity and social justice	2 votes 10 points	3 votes 12 points	5 votes 15 points	3 votes 6 points	3 votes 3 points	46 points
General education	3 votes 15 points	2 votes 8 points	2 votes 6 points	2 votes 4 points	1 vote 1 point	34 points
Tenure track lines	8 votes 40 points	6 votes 24 points	5 votes 15 points	3 votes 6 points	5 votes 5 points	90 points
Family and Medical Leave Act (FMLA)	0 votes 0 points	0 votes 0 points	0 votes 0 points	1 vote 2 points	0 votes 0 points	2 points
Communication	0 votes 0 points	1 vote 4 points	2 votes 6 points	4 votes 8 points	3 votes 3 points	21 points
Classroom issues/technology	3 votes 15 points	5 votes 20 points	4 votes 12 points	5 votes 10 points	6 votes 6 points	63 points

Prioritizing Faculty Issues

1. Enrollment/recruitment and retention

- Brand ambassador
- Building relationships with high schools and feeder colleges
- Language change – Hispanic nurturing rather than “serving”
- Identify activities on campus
- Communicate with parents; help craft potential response content
- Attend NIU graduation
- As an example, the School of Public and Global Affairs brings AP high school students to campus
- Faculty visit prospective students in high schools
- Invest in student organization chapters (perhaps linked to counterparts at high schools)
- On-campus parking concerns
- Review relevant data

2. Tenure track lines

3. Equity gap

4. Decentralizing budget

5. Classroom issues/technology

Return to Table of Contents

Table E - 4

Northern Illinois University

Faculty in Academic Departments by College, Time & Tenure Status

Fall 2014 - Fall 2018

	2014	2015	2016	2017	2018
Business					
Full-Time	89	86	87	88	92
Part-Time	21	26	27	23	27
Total	110	112	114	111	119
Tenured	43	43	44	42	43
Tenure-Track	14	11	15	17	21
No Rank & Other	53	58	55	52	55
Education					
Full-Time	130	129	119	121	120
Part-Time	79	76	85	84	63
Total	209	205	204	205	183
Tenured	62	57	51	52	51
Tenure-Track	20	28	26	30	29
No Rank & Other	127	120	127	123	103
Engineering & Engineering Technology					
Full-Time	46	45	47	52	49
Part-Time	13	9	13	14	13
Total	59	54	60	66	62
Tenured	28	27	27	26	24
Tenure-Track	9	11	13	17	18
No Rank & Other	22	16	20	23	20
Health & Human Sciences					
Full-Time	120	122	123	123	116
Part Time	56	49	54	52	45
Total	176	171	177	175	161
Tenured	38	40	36	36	39
Tenure-Track	32	25	26	26	16
No Rank & Other	106	106	115	113	106

Return to Table of Contents

Table E - 4

Northern Illinois University

Faculty in Academic Departments by College, Time & Tenure Status

Fall 2014 - Fall 2018

	2014	2015	2016	2017	2018
Law					
Full-Time	20	20	18	19	17
Part-Time	5	15	13	8	15
Total	25	35	31	27	32
Tenured	11	8	7	7	6
Tenure-Track	4	7	6	6	6
No Rank & Other	10	20	18	14	20
Liberal Arts & Sciences					
Full-Time	478	460	473	478	457
Part-Time	64	69	65	62	71
Total	542	529	538	540	528
Tenured	240	242	234	239	237
Tenure-Track	67	58	54	48	37
No Rank & Other	235	229	250	253	254
Visual & Performing Arts					
Full-Time	96	95	91	93	90
Part-Time	51	38	36	30	38
Total	147	133	127	123	128
Tenured	64	67	67	68	64
Tenure-Track	11	12	9	9	9
No Rank & Other	72	54	51	46	55
Total Academic Departments					
Full-Time	979	957	958	974	941
Part-Time	289	282	293	273	272
Total	1,268	1,239	1,251	1,247	1,213
Tenured	486	484	466	470	464
Tenure-Track	157	152	149	153	136
No Rank & Other	625	603	636	624	613

Return to Table of Contents

Table A-2A
Northern Illinois University
Total Headcount Enrollment
Fall 1984 - Fall 2018

Official - Tenth Day Count

Fall Semester	Undergrad	Graduate	Law	Total	Total FTE
1984	17,650	5,768	271	23,689	19,313
1985	18,217	5,850	244	24,311	19,868
1986	18,434	6,001	245	24,680	20,154
1987	18,959	6,219	277	25,455	20,720
1988	18,122	5,836	297	24,255	19,824
1989	18,029	6,098	316	24,443	19,892
1990	18,220	5,981	308	24,509	20,051
1991	18,220	6,378	297	24,895	20,123
1992	17,437	6,325	290	24,052	19,375
1993	16,805	6,062	310	23,177	18,531
1994	16,423	6,129	329	22,881	18,131
1995	15,760	6,158	300	22,218	17,460
1996	15,387	5,928	294	21,609	17,013
1997	15,855	5,947	280	22,082	17,518
1998	16,341	5,850	282	22,473	17,962
1999	16,893	5,674	276	22,843	18,364
2000	17,151	5,800	297	23,248	18,603
2001	17,468	6,012	303	23,783	19,103
2002	18,104	6,513	331	24,948	19,958
2003	18,275	6,651	334	25,260	20,348
2004	18,031	6,463	326	24,820	20,092
2005	18,467	6,408	333	25,208	20,501
2006	18,816	6,182	315	25,313	20,758
2007	18,917	6,012	325	25,254	20,630
2008	18,431	5,669	297	24,397	19,947
2009	18,277	5,838	309	24,424	20,022
2010	17,886	5,633	331	23,850	19,592
2011	17,306	5,365	319	22,990	18,817
2012	16,552	4,984	333	21,869	18,033
2013	15,814	5,020	304	21,138	17,324
2014	15,435	4,900	276	20,611	16,940
2015	15,027	4,850	253	20,130	16,502
2016	14,079	4,672	264	19,015	15,712
2017	13,454	4,319	269	18,042	14,982
2018	12,788	4,121	260	17,169	14,352

Note: All full-time equivalents (FTE) are computed as follows:
Total FTE: Undergraduate Credit Hours = 15, Graduate Credit Hours = 12, and
Law Credit Hours = 12

Are GPAs an Inconsistent Measure of College Readiness across High Schools? Examining Assumptions about Grades versus Standardized Test Scores

Elaine M. Allensworth and Kallie Clark
University of Chicago Consortium on School Research
1313 East 60th Street
Chicago, IL 60637

Revised January 27, 2019

Initially Posted April 2018

ABSTRACT

High school GPAs (HSGPAs) are often perceived to represent inconsistent levels of readiness for college across high schools, while test scores (e.g., ACT scores) are seen as comparable. This study tests those assumptions, examining variation across high schools of both HSGPAs and ACT scores as measures of academic readiness for college. We find students with the same HSGPA or the same ACT score graduate at very different rates based on which high school they attended. Yet, the relationship of HSGPAs with college graduation is strong and consistent, and larger than school effects. In contrast, the relationship of ACT scores with college graduation is weak, smaller than high school effects, and the slope of the relationship varies by high school.

High school course grades are critical indicators of academic performance for students, educators, and institutions of higher education. Yet, standardized test scores are often seen as more reliable and objective indicators of academic preparation than students' grades because all students are judged based on the same tasks under the same conditions. All states use standardized tests to judge students' progress toward college readiness goals, with 45 states using ACT or SAT scores (Nayar, 2015). The use of standardized test scores to monitor students' college readiness is recommended clearly in the *What Works Clearinghouse Practice Guide* on how to prepare students for college, while HSGPAs are discussed as one piece of performance data to consider, along with curriculum and assessments (Tierney, Bailey, Constantine, Finkelstein, & Hurd, 2009). A key assumption behind the emphasis on test scores in policy and practice is that college entrance exams are strong and consistent measures of readiness. Yet, the emphasis on test scores over grades in policy and practice recommendations stands in contrast to research showing high school grade point averages (HSGPAs) are stronger predictors than test scores of college outcomes (Bowen, Chingos, & McPherson, 2009; Geiser & Santelices, 2007; Hiss & Franks, 2014; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008).

In this study, we directly address questions about the variability in HSGPAs across high schools as predictors of college readiness, examining whether students with the same HSGPAs are systematically more likely to graduate college if they came from particular high schools, and whether the slope of the relationship differs by high school. We then conduct the same tests with ACT scores. We also discern the extent to which there are high school effects on college graduation that are not captured in either students' HSGPAs or ACT scores.

Prior Literature on the Reliability of Course Grades across Schools and Validity of Tests

Numerous publications give the impression that course grades are not reliable measures of achievement in comparison with test scores. For example, the introduction of a new book on testing and college admissions states:

> ...standardized admissions tests provide a neutral yardstick to assess the performance and promise of students from secondary schools whose course offerings differ sidely in variety and rigor. This is a particularly salient point in an era of widespread grade inflation … (Buckley, Letukas, & Wildavsky, 2018).

Likewise, the introduction of a new report by the Fordham Foundation expresses concern that teachers' grades do not reflect state standards, and wonders how to help parents put more faith in test scores as measures of their students' readiness instead of relying so much on grades (Northern & Petrilli, 2018). These documents reflect current beliefs, which are echoed in the emphasis placed on test scores in policy and in practice recommendations, described above, and often inferred without strong evidence in research studies. However, the evidence is not strong for these beliefs, as described below.

Grades can be seen as non-comparable across schools because they are based on criteria developed by individual teachers, in schools with different curricula. Grades are assigned based on a potentially wide-ranging array of tasks, measured over time, capturing academic knowledge, skills, and academic behaviors, effort, and incorporating teacher judgement. (Bowers, 2011; Brookhart, 1993; Brookhart et al., 2016; Farkas, Sheehan, Grobe, & Shuan, 1990; Kelly, 2008). The fact that grades are based on a wide range of factors, with judgement from many different teachers, makes them potentially highly variable across contexts.

At the same time, the fact that they are based on a large number of raters (teachers) across a wide range of relevant tasks, could actually make them very reliable as indicators of academic readiness for college, where students will also be asked to do a wide range of tasks with different expectations, assessed by many different instructors.

There is no reason to believe *a priori* that tests would necessarily be more reliable than grades as predictors of college performance. Standardized tests assess students on a narrow range of skills (mostly a subset of what students learn in English and math classes) in one type of condition (a timed test), while colleges expect students to have broad knowledge and skills across many subjects, and to show consistent effort in different types of assignments over months at a time. Schools could prepare students for the tests in very different ways (see Koretz, 2017), with different implications for their students' readiness for college.

Moderate correlations with test scores are often used as evidence of unreliability in grades. People sometimes make the argument that grades are "inflated" or "subjective" based on evidence that HSGPAs have increased over time, without concurrent changes in test scores (Camara, Kimmel, Scheuneman, Sawtell, 2004; Gershenson, 2018; Godfrey, 2011; Hurwitz & Lee, 2018), or that students with the same test scores have different HSGPAs at different schools (U.S. Department of Education, 1994; Woodruff & Ziomek, 2004). Pattison, Grodsky, and Muller (2013) describe some of the conceptual flaws in the argument that grades should align tightly with standardized test scores, and suggest focusing instead on the predictive validity of each for later outcomes.

Evidence about the validity of standardized test scores as measures of college readiness has its own weaknesses, making it questionable to use standardized tests as a metric for judging the reliability of grades. SAT and ACT validitation studies tend to be based on improvement in the prediction of college freshman GPAs when test scores are used together with student-reported HSGPA, relative to models that use student-reported HSGPA alone (e.g., Kobrin, et al., 2008; Noble & Sawyer, 2002; Woodruff & Ziomek, 2004). Researchers argue that because the test scores improve the prediction of college freshman GPAs over and above student-reported HSGPAs, they are valid indicators to adjust for inconsistencies in HSGPAs. However, student-reported HSGPAs are more weakly correlated with college freshman GPAs than unweighted HSGPAs taken from transcripts (Geiser & Santelices, 2006; Kuncel, Credé, & Thomas, 2005; Zwick & Himelfarb, 2011); the studies likely over-estimate the value that test scores provide.

Studies based more heavily on HSGPAs from transcripts than student reports suggest test scores provide little improvement in the prediction of college outcomes. Using data from a large sample of colleges across the country, Bowen, Chingos, and McPherson (2009) found the relationship of SAT and ACT scores with college outcomes was small and sometimes not significant (depending on institution type), controlling for HSGPAs, comparing students in the same colleges. In contrast, HSGPAs had a strong relationship with college outcomes controlling for students' test scores. Hiss and Franks (2014) concluded that students in test-optional colleges who did not submit test scores had similar or better college outcomes than students in the same colleges with similar HSGPAs who did submit scores, even though their scores on standardized tests were much lower. Using data from California universities, Rothstein (2004) found that most of the relationship of SAT scores with college GPA could be attributed to high school poverty, school racial composition, and student background.

Grades are lower in harder classes with stronger peers, and this suggests inconsistency in HSGPAs. A number of studies have discerned what are called "frogpond" effects (Attewell, 2001), where students with similar prior test scores, academic performance, or effort receive lower grades in classrooms and schools of predominantly high-achieving students, as compared to those with lower-achieving students (Farkas et al., 1990; Nomi & Allensworth, 2009; Barrow, Sartain, & de la Torre, 2016). Students also tend to get lower grades in classes that are intentionally designed to be challenging, such as Advanced Placement and Honors courses (Sadler & Tai, 2007).

Differences in the types of classes that students take, and the expectations associated with the peer composition, introduce "noise" into the metric of HSGPAs as an indicator of academic performance in high school. ACT and SAT validity studies claim that students' test scores can be used to adjust for different standards and expectations at different schools. There is a need to evaluate that claim using data on HSGPAs from transcripts. It is possible that the overall achievement level in a student's school—information that is publicly available—might be just as useful, or more useful, than individual student's test scores. Two prior studies note that the size of the relationship between HSGPA and college outcomes (graduation or college GPAs) is larger among students within the same high school (i.e., when high school fixed effects are used in a model), than across schools (Bowen et al., 2009; Koretz & Langi, 2018). They interpret this as meaning that HSGPA represents a higher level of achievement at some schools than others, which would be consistent with the "frogpond" effects discussed above, and suggest adjusting for these differences with information on school average achievement.

Different college outcomes for student subgroups with the same HSGPAs have been used as evidence of different grading standards. Other studies have suggested that HSGPAs are inconsistent measures of achievement across high schools because HSGPAs predict that Black and Latino students, and students from low-SES high schools, will receive higher college grades than they actually do (Zwick & Himelfarb, 2011; Zwick, 2013). The researchers suggest that this discrepancy results from differences in the quality of high schools attended by students, and show that school average poverty, used as a proxy for school quality, accounts for some of the differences. They do not conduct equivalent tests of SAT scores to see if similar or larger discrepancies by race or school poverty would occur with prediction models that use SAT scores alone. They also find discrepancies to be much smaller when they use HSGPAs from transcripts, rather than student-reported HSGPAs. Their arguments hold if one is to believe that race, ethnicity, and SES do not affect college success in ways unrelated to acacemic readiness. There are many reasons to believe this is not true, and studies have shown that SAT and ACT scores also overpredict college performance for the same groups (Noble, 2004; Bridgeman, McCamley-Jenkins, & Ervin, 2000; Rothstein, 2004). However, this does suggest that it is important to compare students with similar backgrounds when evaluating the validity of HSGPAs or test scores as indicators of college readiness, and suggests there may be high school effects on students' college outcomes that need to be better understood.

High schools could effect college outcomes in many ways that are not reflected in either students' HSGPA or test performance. For example, high schools might develop structures to prepare students with more "college knowledge" to navigate the post-secondary realm (Hoxby & Turner, 2015; Conley, 2008), or provide a more diverse environment that teaches students to adjust to new situations and people (Tam & Bassett, 2004). Fletcher and Tienda (2007) found that high school fixed effects explained half or more of the differences in college GPA and graduation by students' race and ethnicty, sometimes reversing the relationships. Such high school effects could make it appear that HSGPAs have different value in some schools than others—when there are simply other factors about high schools that also matter for college success.

Contributions of this Study and Research Questions

In this study, we compare the strength and consistency of HSGPAs as predictors of college graduation across high schools with ACT scores, adding to the current literature in a number of ways:

1) **Showing variation across high schools in the relationship of HSGPA with college graduation.** Variation in the predictiveness of HSGPAs by high school could occur either because HSGPAs represent higher levels of readiness from some high schools vs. others (i.e., HSGPAs under- or over-predict graduation for all students at a school), or because the relationship (slope) of the HSGPA differs across high schools (i.e., providing a stronger signal at some schools than others). Prior research has not shown the extent to which there is variation in the relationship of HSGPAs with college graduation by high school.
2) **Conducting equivalent tests on ACT scores as on HSGPA.** Past studies have not explicitly tested whether standardized assessments are comparable across high schools as measures of college readiness. We examine whether students with the same ACT or SAT scores have different college outcomes based on which high school they attended, or if the slope of the relationship of test scores to college outcomes varies by high school.
3) **Discerning the extent to which there are high school effects on college graduation that are not captured in either students' HSGPAs or ACT scores.** While past studies have provided evidence that high school effects on college outcomes exist, they have not quantified the magnitude of high school effects (e.g., the variance across high schools),

that is, how much of a difference it makes which high school a student attended for students who look similar based on their ACT score and HSGPA.

We begin by identifying the extent to which the relationship of each achievement indicator depends on the high school a student attends:

RQ1: How different are college graduation rates for students with the same HSGPAs/ACT scores, who come from different high schools?

We then compare the size and consistency of the relationships of HSGPAs and ACT scores with college graduation, and examine whether including students' ACT scores in the prediction of college graduation substantially reduces inconsistency across high schools over using HSGPA alone:

RQ2: Do ACT scores provide a stronger, or more consistent, prediction of college readiness across high schools than HSGPAs?

RQ3: Is there less high school variance in college graduation rates in models that use students' ACT scores and HSGPAs together, than models that use HSGPAs alone?

Finally, we show the extent to which information about high schools (school poverty and average ACT scores) explain high-school level variation.

RQ4: To what extent are high school differences in college graduation rates for students with the same HSGPAs and ACT scores explained by school achievement level and school poverty?

Research Methods

This study uses data from the Chicago Public Schools (CPS), a large, public school district that contains schools with varying academic composition—extremely high-achieving selective schools that get ranked among the top high schools in the country, heterogeneous schools, and schools with very low test scores. We include for analysis all students who graduated from neighborhood, magnet, selective, and vocational high schools between the years of 2006 and 2009, who enrolled in a four-year college immediately following graduation, and who had complete data (n=17,753).[1] Table 1 provides summary statistics of the analytic group and variables used in the models.

We only include students who enrolled in a four-year college, so as not to confound enrollment in college with ability to succeed in college once enrolled. Because college admissions use HSGPA and ACT scores to determine who is accepted, those measures will be related to college graduation simply because they provide access, regardless of whether they indicate readiness to succeed once enrolled. By comparing only students who enrolled in college, and controlling for institutional characteristics (described below), we focus on the extent to which the HSGPAs and ACT scores are indicators of students' likelihood of succeeding once in college, not the degree to which they are signals to admissions officers.

Data and Variables

Data on academic performance and student demographic information (gender, race, and ethnicity) come from district administrative datasets. We obtained economic information on students' residential neighborhoods by linking students' addresses to information from the U.S. Census at the level of block groups on the percent of adult males employed and the percent of families with incomes above the poverty line. HSGPAs were created by coding grades in

[1] Charter school graduates were not included because their transcripts are not available. A total of 2,595 cases had missing data: 934 were missing cumulative HSGPAs, 982 were missing ACT scores, and 828 were missing the IPEDS institutional graduation rate for the college they attended. Students in the restricted group had slightly higher HSGPAs (2.72 vs 2.69) and ACT scores (20.12 vs 19.97) than students in the total population. The groups were nearly identical with regards to ethnicity, race, gender, SES, and institutional graduation rate.

students' transcripts 0 through 4 (F through A), and creating an unweighted average of all courses completed in high school. At the time the students were in high school, all students in Illinois took the ACT during the spring of the eleventh grade. College enrollment records and six-year graduation outcomes were obtained through the National Student Clearinghouse. Students were included for analysis if they had full-time enrollment records in a four-year college during the fall term after they graduated high school. Six-year college graduation is defined as earning a four-year college degree within six years of graduating from high school. Appendix Table A1 shows college graduation rates by students' ACT scores and HSGPA unadjusted for demographic and college characteristics or high school effects.

Colleges offer different supports and structures which influence whether students graduate (Bowenet al., 2009; Cohodes & Goodman, 2012; Kurlaender & Grodsky, 2013). Therefore, it was important to control for college characteristics. We did this by including information on colleges obtained through the Integrated Postsecondary Education Data System (IPEDS) as covariates: the race- or ethnicity-specific six-year institutional graduation rate of the college (matched to the race and ethnicity of the student), college size (number of undergraduate students), the percentage of freshman students who are full time, and the student-to-faculty ratio.

Methods

We estimated the variance in college graduation rates by high school using hierarchical linear models, with students nested within high schools. We considered using cross-nested models with students simultaneously nested within their high school and college, or controlling for college fixed effects. However, students in our analysis group matriculated to more than 500 different four-year colleges across the U.S., and at many of these colleges there were only a small number of students. This resulted in imprecise estimates of college effects for a large proportion of the sample through these other methods.

For RQ1, we used two different methods of estimating variation in college graduation rates by high school for students with the same HSGPA/ACT score. First, we used a series of dummy variables to model the relationship between HSGPA and college graduation non-parametrically. There is no intercept, so the coefficient for each HSGPA dummy variable represents the average college graduation rate for students in that HSGPA group. We allowed the coefficients to vary by high school to identify the variation in college graduation rates across high schools for students with the same HSGPAs. These same models were then repeated with student ACT bins in lieu of HSGPA bins. Coefficients for other covariates were fixed across schools, predicting the log odds of graduating from a four-year college in six years:

Level-1 Model [1]

$$\log (p_{grad}/1-p_{grad})_{ij} = \sum_{s=1}^{5} \beta_{sj} (S)_{ij} + \sum_{g=6}^{21} \beta_{gj} (G)_{ij} + \sum_{c=22}^{25} \beta_{cj} (C)_{ij} + r_{ij}$$

Level-2 Model

$$\beta_{sj} = \gamma_{s0}$$
$$\beta_{gj} = \gamma_{g0} + u_{gj}$$
$$\beta_{cj} = \gamma_{c0}$$

S is a vector of student background variables (Neighborhood poverty, male, Black, Latino, and Asian).

G is a vector of dummy variables representing HSGPA bands.

C is a vector of college institutional variables.

u_{gj} is the high school-level variance in college graduation rates for students in the HSGPA band, controlling for student background and college institutional variables.

The above method assumes no particular functional form. However, because students with different levels of achievement are not evenly distributed across schools, not all high schools have students in all achievement bands. Therefore, we only calculated school-level random effects for bands in which at least 95 percent of schools are represented.

We also ran models which use standardized continuous versions of HSGPAs, rather than the binned variables, and calculated the average school effect across all achievement levels. These models include a squared term, since the relationship of each achievement measure is slightly quadratic. We ran models in which the slopes of each achievement measure with college graduation are fixed, and models that allow the slopes of the relationships to vary by high school. As the results are similar, only the second are shown in the manuscript, and the first are available from the authors:

Level-1 Model [2]

$$\log (p_{grad}/1\text{-}p_{grad})_{ij} = \beta_{0j} + \sum_{s=1}^{5} \beta_{sj} (S)_{ij} + \beta_{6j}(ZGPA)_{ij} + \beta_{7j}(ZGPA^2)_{ij} + \sum_{c=8}^{11} \beta_{cj} (C)_{ij} + r_{ij}$$

Level-2 Model

$$B_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{sj} = \gamma_{s0}$$
$$\beta_{6j} = \gamma_{60} + u_{6j}$$
$$\beta_{7j} = \gamma_{70} + u_{7j}$$
$$\beta_{cj} = \gamma_{c0}$$

In Equation 2, u_{0j} is the high school-level variance in college graduation rates, controlling for students' HSGPA, student background variables and the institutional characteristics of the colleges in which they enroll. Variance components on the slopes, u_{6j} and u_{7j}, show variation in the size the relationship of HSGPA with college graduation across high schools--whether grades are stronger measures of college readiness at some schools than others. Equations 2 was replicated with ACT scores.

Finally, we ran models that entered HSGPA and ACT scores together in the models to discern how much ACT scores improve the prediction of college graduation beyond using HSGPAs alone, to answer question 3:

Level-1 Model, Equation 2 [3]

$$\log (p_{grad}/1\text{-}p_{grad})_{ij} = \beta_{0j} + \sum_{s=1}^{5} \beta_{sj} (S)_{ij} + \beta_{6j}(ZGPA)_{ij} + \beta_{7j}(ZACT)_{ij} + \beta_{8j}(ZGPA^2)_{ij} + \beta_{9j}(ZACT^2)_{ij} + \sum_{c=10}^{13} \beta_{cj} (C)_{ij} + r_{ij}$$

Level-2 Models

$$B_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{sj} = \gamma_{s0}$$
$$\beta_{6j} = \gamma_{60} + u_{6j}$$
$$\beta_{7j} = \gamma_{70} + u_{7j}$$
$$\beta_{6j} = \gamma_{80} + u_{8j}$$
$$\beta_{7j} = \gamma_{90} + u_{9j}$$

$$\beta_{cj} = \gamma_{c0}$$

We did this in two ways. First, we grand-mean centered all student variables to show the overall relationships, and then we group-mean centered the variables to discern the relationship of each with college graduation relative only to other students in the same school. This second specification is similar to a school fixed-effects model. To address RQ4, we included school-level predictors of school performance level (average ACT score) and school poverty as predictors of B_{0j}.

Results

Table 2 displays coefficients from models predicting college graduation rates with HSGPA, without and with covariates. The odds ratios show the likelihood of graduating from college; students with a 3.0-3.25 HSGPA have fairly even odds (0.91), which gives them just under a 50-50 chance (48 percent probability), while students with a HSGPA of 3.5-3.75 are 3.6 times more likely to graduate as to not graduate (odds of 3.65, or about 78 percent graduating and 22 percent not graduating). HSGPA has a strong relationship with college graduation in both the unconditional model and the model that controls for students' backgrounds and college institutional variables, although the relationship is smaller once the control variables are introduced. The coefficients from the full model are converted into percentages and displayed graphically as the thick black line in the left panel of Figure 1. Across the range of HSGPAs, the probability of graduating from college ranges from 20 percent for students with HSGPAs less than 1.5 to about 80 percent for students with HSGPAs of 3.75 or higher, after controlling for student backgrounds and college characteristics.

The random effects at the bottom of Table 2 show the degree to which average graduation rates vary across high schools among students in each HSGPA bin. There is significant high school variance in college graduation rates for students in each HSGPA bin. For example, among students with HSGPAs between 3.25-3.5, a two-standard deviation range of high school effects is 0.144 ± 0.575 in log-odds in the conditional model. Thus, students with a 3.25-3.5 HSGPA at schools with very negative school effects (one standard deviation below the mean) have college graduation rates that are similar to students with HSGPAs of 2.75-3.0 at more typical schools (where the odds of graduating are 0.72).

Model 2 in Table 2 shows the results from a model where HSGPA is entered as a continuous variable along with a squared term, instead of discrete bins. The linear component shows that for every standard deviation increase in HSGPA, the odds of graduating from college double (odds coefficient = 2.02) at the point where the quadratic term is zero (which is at the sample average). The quadratic term is positive, so the relationship is larger among students with the highest levels of achievement, and lower among students with low HSGPAs. The school variance component for the intercept from this model (0.603) is slightly higher than those in the binned model (where variance components ranged from 0.501 to 0.575), and represents the variance in school effects averaged across students of all achievement levels. Not only is the school-level variance component large (0.603), it is larger when HSGPAs are included in the model than in a model that only includes control variables (0.447, not shown in table). This pattern is consistent with the “frogpond” effects discussed earlier, wherein HSGPAs are suppressed at high schools with more positive school effects. About one-fourth of the school-level variation in Model 2 ((0.603-0.447)/0.603 = 26%) is “extra” variation that is induced by comparing students with similar HSGPAs.

The model displayed in Table 2 also allows the slope of the relationship between HSGPA and college graduation to vary by high school. The strong linear trend (coefficient of 0.703), does not vary significantly by high school. The quadratic term (coefficient of 0.062) does vary slightly across schools (0.103). The noise that is introduced by variation in the linear and quadratic components is small relative to the signal from the linear slope (0.703), so the overall slope of the relationship is fairly similar across schools. The gray lines in the left panel of Figure 2 show the relationship of HSGPA with college graduation for each high school, estimated from the coefficients and variance components from Model 2. The considerable variation in college graduation rates by high school for students with the same HSGPA is clearly visible. At the same time, the relationship between HSGPA and college graduation has a similar slope, and is large and positive, across high schools.

Table 3 shows the results of models that mirror those in Table 2, substituting ACT scores for HSGPAs. Differences in college graduation rates by ACT score are more modest than by HSGPA, particularly after controlling for student background and college characteristics, but show a sizable range—from odds of 0.39 to 1.98 in the conditional model (graduation rates of 28 to 66 percent). School-level variance is smaller among students with the same ACT score than among students with the same HSGPA. Still, there is considerable variation in college graduation rates by high school among students with the same ACT score (0.265 to 0.343). For students with an ACT score of 16-17, for example, a two-standard deviation range in the log-odds of graduating is -0.387 ± 0.343. Students with an ACT score of 16-17 in a school with large positive effects (one standard deviation above the mean) would graduate at a rate that similar to students with scores of 20-21 in a more typical school. Thus, students with the same qualifications, defined by either their HSGPA or their ACT score, graduate at different rates based upon which high school they attend.

Model 2 in Table 3 shows the relationship of ACT scores with college graduation modeled with continuous linear and quadratic terms. The standardized linear term is much smaller than that of standardized HSGPA scores (0.129 vs. 0.703), with the odds of graduating increasing by 14 percent (odds coefficient of 1.14) for every standard deviation increase in ACT scores when the quadratic term equals zero. There is a negative quadratic term, so the relationship is larger among students with low achievement, small among students with high achievement, and becomes negative among students with the highest achievement. The variance components show that the linear component of the slope varies significantly, and the variance in the slopes (0.192) is larger than the average slope (0.129). Thus, the noise introduced by school effects is larger than the signal from ACT scores. Where students attend high school says more about whether they are likely to graduate from college than their individual ACT score, at least among students with average or high ACT scores.

ACT scores also provide less accurate predictions of college success based on students' race, ethnicity, and gender than HSGPAs. The subgroup differences in college graduation rates are significantly different from zero for Asian and male students in the models that control for ACT scores, but the demographic coefficients are not significantly different from zero in the models that control for HSGPAs. ACT scores explain only a little of the school-level variance in college graduation rates; the variance component on average school effects (0.411) is similar to a model with the same control variables but no ACT scores (0.446). However, they do not induce more school-level variance, as was seen with HSGPAs.

The right panel of Figure 1 shows the relationships from Models 2 and 3, modeled as percentages. The dark line shows the averages from the bins in Model 2, while the gray lines show the relationship for each school, calculated from the coefficients and variance components in Model 2. The dark line is not at the center of the gray lines because most of the students with high ACT scores are concentrated in schools with high average college graduation rates, while students with very low ACT scores are concentrated at schools with low average college graduation rates. Many schools do not have students with very high ACT scores, and a number of other schools do not have students with very low ACT scores, so few of the lines go the full range of the horizontal axis. The figure shows how the relationship of students' individual ACT scores with college graduation is small relative to the variation in.

In Table 4, ACT scores and HSGPAs are included together in the models. The main HSGPA coefficient does not change substantially relative to the model without ACT scores in Table 2 (0.708 vs 0.703), but the main ACT coefficient shrinks considerably from the model without HSGPA (from 0.129 to a nonsignificant -0.016). Because the ACT score contributes little to the prediction, there is a similar amount of school-level variance in the combined model (0.622) as the model that includes HSGPA alone (0.603, from Table 1). ACT scores used at the individual student level do not reduce the variability by high school in predicting who will graduate college. The slope of the relationship of ACT scores with college graduation still varies significantly based on high school (0.213); in schools a standard deviation below the mean the linear slope is negative (-0.016 - .213, or -0.219) and in others it is positive (-0.016 + 0.213 or

0.197).

In the next model, the variables are group-mean centered so that the coefficients show the relationship of each variable with college graduation relative to other students in the same school. The school-level variance of the intercept in this model is much larger because the student variables do not control for differences across schools in student body composition. The within-school coefficient for HSGPAs is slightly larger than the coefficient from the earlier model, while the ACT score coefficient is small and not significant. The ACT slope varies significantly by high school (0.206, $p<0.000$) while the main linear portion of GPA slope does not vary and the quadratic term varies only slightly.

In final model, we include predictors of school performance level (average ACT scores among all students) and school poverty level. School average ACT scores are significantly related to college graduation, explaining school-level differences among students with the same HSGPAs and individual ACT scores. The odds of graduating college increase by 60 percent for every standard deviation increase in school average ACT scores, for students with the same HSGPA and ACT score. Average ACT scores in the school reduce the high school variation in college graduation rates by 42 percent (0.324 vs. 0.622). The school poverty level is not significant in this model, but that is because it is highly correlated with school average ACT scores ($r=0.70$). If entered alone in the model, either variable is a significant predictor with odds ratios of 0.70 for school poverty and 1.68 for school average ACT.

Discussion

It is commonly believed that HSGPAs indicate different levels of readiness for college, based on the high school a student attended, while ACT scores are consistent indicators.

However, HSGPAs perform in a strong and consistent way across high schools as measures of college readiness, while ACT scores do not. There are large high school effects on college graduation, such that students with either the same HSGPA or the same ACT score graduate from college at different rates, based on which high school they attended. Neither capture all of the ways in which high schools influence college graduation. The school differences are larger for students with the same HSGPA, which is consistent with prior studies showing that grades are depressed in schools and classes with higher-achieving students. HSGPAs are not equivalent measures of readiness across high schools, but they are strongly predictive in all schools, and the signal they provide is larger than the differences across schools. School-level variance in college graduation rates is one-quarter smaller among students with the same ACT score than students with the same HSGPA. However, this still leaves considerable school-level variance, and the signal provided by ACT scores is much smaller than the noise introduced by school effects.

As measures of individual students' academic readiness, ACT scores show weak relationships, and even negative relationships at the higher achievement levels. The negative slope among students with the highest achievement could result if people are using ACT scores to make decisions about students' readiness for very rigorous academic programs out of a belief that they are strong indicators of readiness, when they are not. Future research might investigate this further. Regardless, there is little evidence that students will have more college success if they work to improve their ACT score, as most of the signal from the ACT score seems to represent factors associated with the student's school, rather than the student. In contrast, students' efforts to improve their HSGPAs would seem to have considerable potential leverage for improving college readiness. The fact that HSGPAs are based on so many different criteria—including effort over an entire semester in many different types of classes, demonstration of skills through multiple formats, and different teacher expectations—does not seem to be a weakness. Instead, it might help to make HSGPAs strong indicators of readiness, since they measure a very wide variety of the skills and behaviors that are needed for success in college, where students will also encounter widely varying content and expectations.

Test scores provide more of a signal at the school level, with school-level average test scores providing additional information about students' likelihood of graduating above and beyond students' individual HSGPAs. For judging college readiness (e.g., college admissions), school-average ACT scores would provide a stronger prediction than students' individual scores. This is consistent with the findings and recommendations in Koretz and Langi (2018) and Bowen, Chingos, and McPherson (2009). The same pattern is observed with school-average poverty levels (in models that do not control for average ACT scores), which echoes Rothstein's (2004) findings. High school effects could result from higher academic standards (e.g., more college-oriented curricula at higher-achieving, higher-SES schools). Yet, they could also represent selection effects. Families with more financial, social, and human capital might select into higher-achieving, higher-SES high schools, either by choice of residence or application, and those families would likely continue to offer financial support when students are in college. School effects also could come from different peer networks, advising, supplemental experiences, or broader curricular offerings available at schools with more resources. Future research should investigate high school effects on college outcomes more thoroughly.

This study was conducted only with data from Chicago, and only with data from public schools. There could be more variation across high schools with a more comprehensive sample, and different relationships. The similarity in results that are available from studies of schools in other places provide some indication of their generalizability. Studies that use data from samples that include 21 prestigious flagship universities from across the country and all public universities in four states (Bowen et al., 2008; Koretz & Langi, 2018; Rothstein, 2004) all show that HSGPAs are strongly related to either college graduation or to college freshman GPA, and that students' individual ACT or SAT scores add only modestly to the prediction beyond HSGPA, if at all, in models that include high school fixed-effects. The graduation rates presented by Bowen, Chingos, and McPherson (2008) for specific HSGPAs are also similar to the graduation rates found here and shown in Figure 1. Graduation rates by HSGPA are not provided in other studies, to our knowledge.

This research strongly supports the use of students' grades in a formative way, to guide school improvement efforts and assess the effectiveness of programs designed to improve college readiness, and relying much less heavily on test scores. The teachers and schools that improve test scores are not always the same as those that improve students' grades (Jackson, 2016), and programs that have positive effects on test scores do not always have positive effects on grades (Nomi & Allensworth, 2009). Reaching goals that all students will graduate college-ready would seem to require strategies around improving students' HSGPAs, since HSGPAs are so strongly related to eventual college completion at all high schools. Higher ACT scores might help students get access to stronger colleges, but the pay-off would only occur if students actually attend stronger colleges. As an increasing number of colleges become test-optional, they are likely to be decreasingly salient for college admissions, as well.

States and districts might also consider relying less heavily on standardized test scores in their accountability systems as indicators of college readiness, given that the relationship is not strong and not consistent across schools. A number of states have developed longitudinal data systems that allow for the creation of metrics of students' actual performance in college. The existence of large school effects among students with the same ACT scores suggests that if high schools are not tracking the success of their students in college, and are relying solely on students' test scores as indicators of their students' college readiness, they may be misestimating the effects of their practices on students' college readiness. Likewise, we worry that if families and college admissions officers must rely on school poverty levels and average test scores as proxy indicators for school effects, they might not recognize strong practices at schools serving low-income students. Measuring and publishing school effects on postsecondary outcomes would provide better information to guide families, educators, and policymakers.

References

Attewell, P. (2001). The winner-take-all high school: Organizational adaptations to educational stratification. *Sociology of Education 74*(4), 267-295.

Barrow, L., Sartain, L., & de la Torre, M. (2016). *The role of selective high schools in equalizing educational outcomes: Heterogeneous effects by neighborhood socioeconomic status* (Working Paper No. 2016-17). Chicago, IL: Federal Reserve Bank of Chicago.

Bowen, W.G., Chingos, M.M., & McPherson, M.S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT® I: Reasoning test*. New York, NY: College Entrance Examination Board.

Bowers, A.J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation*, *17*(3), 141-159.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, *86*(4), 803-848.

Buckley, J., Letukas, L., & Wildavsky, B., (Eds). (2018). *Measuring success: Testing, grades, and the future of college admissions*. Baltimore, MD: Johns Hopkins University Press.

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E.A. (2004). *Whose grades are inflated?* (Research Report No. 2003-4). New York, NY: College Entrance Examination Board.

Cohodes, S., & Goodman, J. (2012). *First degree earns: The impact of college quality on college completion rates* (HKS Faculty Research Working Paper Series RWP12-033). Cambridge, MA: John F. Kennedy School of Government, Harvard University.

Conley, D. T. (2008). Rethinking college readiness. *New directions for higher education*, *2008*(144), 3-13.

Farkas, G., Sheehan, D., & Grobe, R.P. (1990). Coursework mastery and school success: Gender, ethnicity, and poverty groups within an urban school district. *American Educational Research Journal*, *27*(4), 807-827.

Fletcher, J., & Tienda, M. (2010). Race and ethnic differences in college achievement: Does high school attended matter?. *The Annals of the American Academy of Political and Social Science*, *627*(1), 144-166.

Geiser, S., & Santelices, M.V. (2006). The role of Advanced Placement and honors courses in college admissions. In P. Gandara, G. Orfield, & C. Horn (Eds.), *Expanding opportunity in higher education* (pp. 75-114). Albany, NY: SUNY Press.

Geiser, S., & Santelices, M.V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. Berkeley, CA: Center for Studies in Higher Education.

Gershenson, S. (2018). *Grade inflation in high schools (2005–2016).* Washington, DC: Thomas B. Fordham Institute.

Godfrey, K.E. (2011). *Investigating grade inflation and non-equivalence* (Research Report 2011-2) New York, NY: College Entrance Examination Board..

Hiss, W.C., & Franks, V.W. (2014). *Defining promise: Optional standardized testing policies in American college and university admissions*. Arlington, VA: The National Association for College Admission Counseling.

Hoxby, C.M., & Turner, S. (2015). What high-achieving low-income students know about college. *The American Economic Review*, *105*(5), 514-517.

Hurwitz, M., & Lee, J. (2018). Grade inflation and the role of standardized testing. In J. Buckley, L. Letukas, & B. Wildavsky (Eds.) *Measuring success: Testing, grades, and the future of college admissions*. Baltimore, MD: Johns Hopkins University Press.

Jackson, C.K. (2016). *What do test scores miss? The importance of teacher effects on non-test score outcomes* (No. w22226). Cambridge, MA: National Bureau of Economic Research. Retrieved from https://www.nber.org/papers/w22226

Kelly, S. (2008). What types of students' effort are rewarded with high marks?. *Sociology of Education*, *81*(1), 32-52.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT® for predicting first-year college grade point average (Research Report No. 2008-5). New York, NY: College Entrance Examination Board.

Koretz, D., & Langi, M. (2018). Predicting freshman grade-point average from test scores: Effects of variation within and between high schools. *Educational Measurement: Issues and Practice*, *37*(2), 9-19.

Koretz, D. (2017). *The Testing Charade*. Chicago: University of Chicago Press.

Kuncel, N.R., Credé, M., & Thomas, L.L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75*(1), 63-82.

Kurlaender, M., & Grodsky, E. (2013). Mismatch and the paternalistic justification for selective college admissions. *Sociology of Education*, *86*(4), 294-310.

Nayar, N. (2015). *How are States reporting on college and career readiness?* Washington, DC: College & Career Readiness & Success Center at American Institutes for Research.

Noble, J. (2004). The effects of using ACT composite scores and high school averages on college admissions decisions for ethnic groups. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 303-319). New York, NY: Routledge.

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT composite score*. Iowa City, IA: ACT, Inc.

Nomi, T., & Allensworth, E. (2009). “Double-dose” algebra as an alternative strategy to remediation: Effects on students' academic outcomes. *Journal of Research on Educational Effectiveness*, *2*(2), 111-148.

Northern, A.M., & Petrilli, M.J. (2018). Foreword and executive summary. In S. Gersheson (Ed.), *Grade inflation in high schools (2005-2016)*. Washington, DC: Thomas B. Fordham Institute.

Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, *42*(5), 259-265.

Rothstein, J.M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, *121*(1), 297-317.

Sadler, P.M., & Tai, R.H. (2007). Weighting for recognition: Accounting for advanced placement and honors courses when calculating high school grade point average. *NASSP Bulletin*, *91*(1), 5-32.

Tam, M.Y.S., & Bassett, G.W. (2004). Does diversity matter? Measuring the impact of high school diversity on freshman GPA. *Policy Studies Journal*, *32*(1), 129-143.

Tierney, W. G., Bailey, T., Constantine, J., Finkelstein, N., & Hurd, N. F. (2009). *Helping students navigate the path to college: What high schools can do: A practice guide (NCEE #2009-4066)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

U.S. Department of Education, Office of Educational Research and Improvement. (1994). *What do student grades mean? Differences across schools* (Office of Research Report 94-3401). Washington, DC: U.S. Department of Education.

Woodruff, D.J., & Ziomek, R.L. (2004). *Differential grading standards among high schools* (ACT Research Report Series, 2004-02). Iowa City, IA: ACT, Inc.

Zwick, R., & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement*, *48*(2), 101-121.

Zwick, R. (2013). Disentangling the role of high school grades, SAT® scores, and SES in predicting college achievement. *ETS Research Report Series*, *2013*(1), 1-20.

Tables and Figures

Table 1. Descriptive Statistics of Variables Used in the Models

		Mean	Std. Dev.
Demographic Characteristics	Male	37%	
	Black	50%	
	Latino	26%	
	Asian	10%	
	White	14%	
	Neighborhood Poverty (standardized across all students, not just college-goers)	-0.12	0.99
High School Achievement	Cumulative HSGPA	2.72	0.65
	ACT Composite Score	20.12	4.33
College Outcome	College Degree in Six Years	49%	
College Institutional Characteristics	College Size (# Freshmen)	3662	2390
	% Full Time Freshmen	65%	17%
	Student to Faculty Ratio	17	5.43
	Six-Year Institutional Graduation Rate for student's racial or ethnic group	47%	22%

Based on students who enrolled in a four-year college the fall after graduation (n=17,753). Institutional characteristics are based on the college freshmen cohort of 2008.

Table 2. Model Predicting Six-Year College Graduation Rates by Student HSGPA Score
Students Nested within High School

Coefficients	Unconditional GPA Binned			Model 1 GPA Binned			Model 2 Random GPA slope		
	Coeff	s.e.	odds	Coeff	s.e.	odds	Coeff	s.e.	odds
Male				-0.096	0.04	0.91	-0.092	0.04	0.91
Black				-0.024	0.07	0.98	0.127	0.08	1.14
Latino				-0.077	0.06	0.93	0.013	0.07	1.01
Asian				0.052	0.08	1.05	0.046	0.08	1.05
ZPoverty				-0.093	0.02	0.91	-0.069	0.02	0.93
ZCollege Size				0.061	0.02	1.06	0.016	0.00	1.02
Z%Full Time Students				0.013	0.03	1.01	0.000	0.00	1.00
ZStudent-Faculty Ratio				0.151	0.03	1.16	0.833	0.15	2.30
ZCollege Grad Rate				0.487	0.03	1.62	0.019	0.00	1.02
GPA <1.5	-1.834	0.11	0.16	-1.410	0.11	0.24			
GPA 1.5-1.75	-1.720	0.11	0.18	-1.319	0.11	0.27			
GPA 1.75-2.0	-1.210	0.07	0.30	-0.875	0.07	0.42			
GPA 2.0-2.25	-1.429	0.10	0.24	-1.025	0.09	0.36			
GPA 2.25-2.5	-1.185	0.10	0.31	-0.844	0.09	0.43			
GPA 2.5-2.75	-0.834	0.09	0.43	-0.593	0.08	0.55			
GPA 2.75-3.0	-0.496	0.09	0.61	-0.326	0.08	0.72			
GPA 3.0-3.25	-0.098	0.10	0.91	-0.002	0.09	1.00			
GPA 3.25-3.5	0.144	0.10	1.15	0.144	0.08	1.15			
GPA 3.5-3.75	1.296	0.07	3.65	0.939	0.07	2.56			
GPA 3.75-4.0	1.830	0.11	6.23	1.320	0.11	3.74			
ZGPA							**0.703**	**0.03**	**2.02**
$ZGPA^2$							0.062	0.02	1.06
Intercept							-0.558	0.07	0.57
Variance Components	**Variance of Coefficients across High Schools** *In Standard Deviations*								
	s.d.	p-value		s.d.	p-value		s.d.	p-value	
GPA 2.25-2.5	0.792	0.000	***	0.522	0.000	***			
GPA 2.5-2.75	0.745	0.000	***	0.598	0.000	***			
GPA 2.75-3.0	0.754	0.000	***	0.500	0.000	***			
GPA 3.0-3.25	0.863	0.000	***	0.501	0.000	***			
GPA 3.25-3.5	0.814	0.000	***	0.575	0.000	***			
ZGPA							0.107	0.106	
$ZGPA^2$							0.103	0.032	*
Intercept							0.603	0.000	***

*p<0.05, **p<0.01, ***p<0.001. Student background and college institutional control variables were grand-mean centered in all models. Variables beginning with Z were standardized, except squared terms which are the square of the standardized variables. A model with only the control variables, without HSGPA, produces a school-level variance component of 0.447 in standard deviation units.

Table 3. Model Predicting 6-Year College Graduation Rates by Student ACT Score
Students Nested within High School

Coefficients	Unconditional ACT Binned			Model 1 ACT Binned			Model 2 Random ACT slope		
	Coeff	s.e.	odds	Coeff	s.e.	odds	Coeff	s.e.	odds
Male				-0.346	0.04	0.71	-0.342	0.04	0.71
Black				-0.045	0.08	0.96	0.026	0.08	1.03
Latino				-0.056	0.06	0.95	-0.012	0.06	0.99
Asian				0.231	0.10	1.26	0.227	0.08	1.25
ZPoverty				-0.091	0.02	0.91	-0.069	0.02	0.93
ZCollege Size				0.005	0.02	1.00	0.009	0.02	1.01
Z%Full Time Students				0.059	0.02	1.06	0.058	0.02	1.06
ZStudent-Faculty Ratio				0.144	0.03	1.15	0.139	0.02	1.15
ZCollege Grad Rate				0.673	0.04	1.96	0.671	0.03	1.96
ACT < 14	-1.59	0.11	0.20	-0.941	0.11	0.39			
ACT14-15	-1.01	0.07	0.37	-0.482	0.07	0.62			
ACT16-17	-0.793	0.06	0.45	-0.387	0.06	0.68			
ACT18-19	-0.489	0.06	0.61	-0.231	0.05	0.79			
ACT20-21	-0.012	0.07	0.99	-0.059	0.07	0.94			
ACT22-23	0.552	0.09	1.74	0.309	0.09	1.36			
ACT24-25	0.852	0.08	2.34	0.407	0.08	1.50			
ACT26-27	0.986	0.11	2.68	0.356	0.10	1.43			
ACT28-29	1.46	0.15	4.33	0.684	0.15	1.98			
ACT30+	1.58	0.17	4.86	0.506	0.18	1.66			
ZACT							**0.129**	**0.04**	**1.14**
$ZACT^2$							-0.099	0.02	0.91
Intercept							-0.251	0.06	0.78
Variance Components	**Variance of Coefficients across High Schools** *In Standard Deviations*								
	s.d.	p-value		s.d.	p-value		s.d.	p-value	
ACT14-15	0.446	.002 **		0.343	.040*				
ACT16-17	0.447	.000***		0.343	.000***				
ACT18-19	0.402	.000***		0.265	.002**				
ZACT							0.192	.012 *	
$ZACT^2$							0.067	.424	
Intercept							0.411	.000 ***	

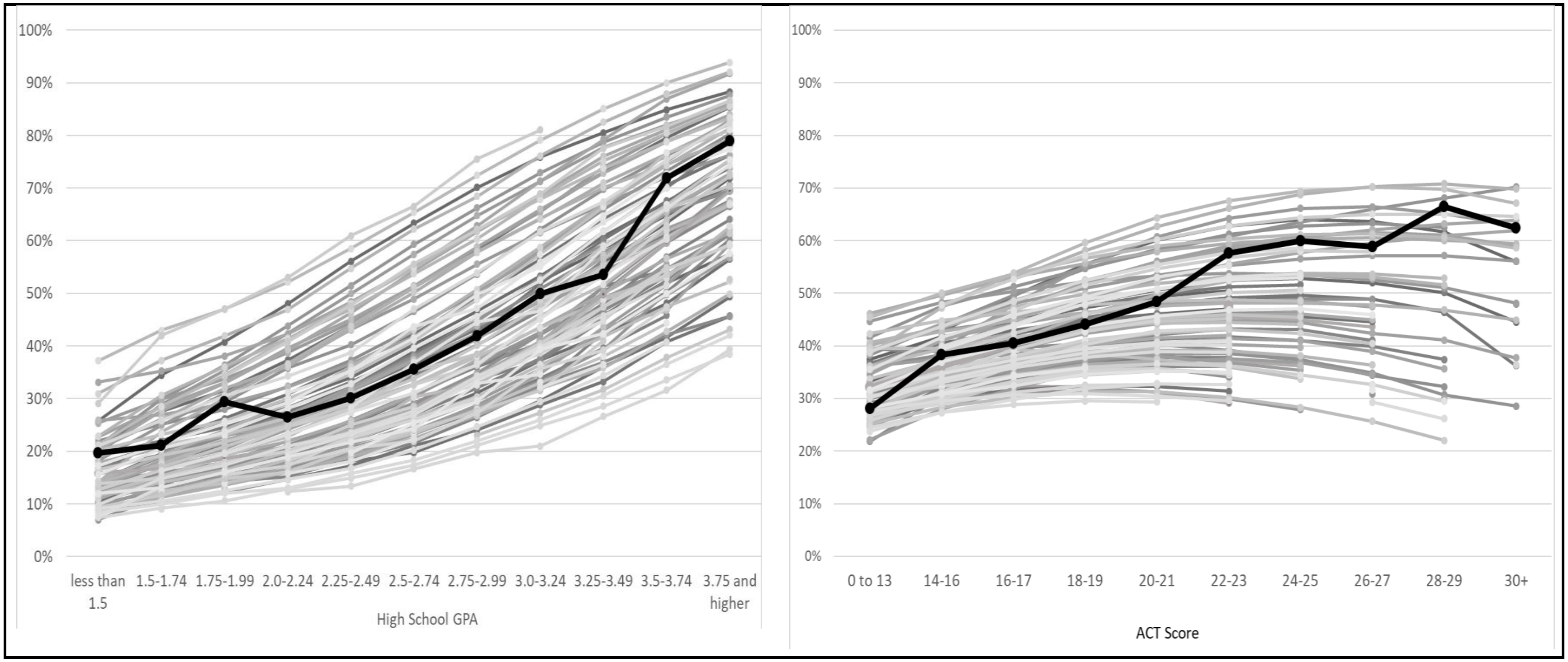
***p<0.001, **p<0.01, *p<0.05. Student background and college institutional control variables were grand-mean centered in all models. Variables beginning with Z were standardized, except squared terms which are the square of the standardized variables. A model with only the control variables, without ACT scores, produces a school-level variance component of 0.447 in standard deviation units.

Table 4. Models Predicting 6-Year College Graduation Rates by Both HSGPA and ACT Score
Students Nested within High School

	Varying Slopes			Group-Mean Centered (School Fixed Effects)			School-Level Variables		
	Coeff	s.e.	odds	Coeff	s.e.	odds	Coeff	s.e.	odds
Male	-0.086	0.04	0.92	-0.08	0.04	0.92	-0.08	0.04	0.92
Black	0.134	0.08	1.14	0.17	0.08	1.19	0.17	0.06	1.18
Latino	0.011	0.07	1.01	0.02	0.07	1.02	0.00	0.07	1.00
Asian	0.042	0.08	1.04	0.03	0.08	1.03	0.04	0.10	1.04
ZPoverty	-0.071	0.02	0.93	-0.06	0.02	0.94	-0.06	0.02	0.94
ZCollege Size	0.089	0.02	1.09	0.09	0.02	1.10	0.10	0.02	1.10
Z%Full Time Students	-0.006	0.03	0.99	-0.01	0.03	0.99	-0.01	0.02	0.99
ZStudent-Faculty Ratio	0.151	0.03	1.16	0.15	0.03	1.16	0.16	0.03	1.17
ZCollege Grad Rate	0.422	0.04	1.52	0.42	0.04	1.51	0.42	0.04	1.52
ZGPA	0.708	0.03	2.03	0.73	0.03	2.07	0.75	0.03	2.12
$ZGPA^2$	0.063	0.02	1.06	0.05	0.02	1.05	0.06	0.02	1.06
ZACT	-0.016	0.04	0.98	-0.07	0.04	0.94	-0.02	0.04	0.98
$ZACT^2$	-0.108	0.02	0.90	-0.10	0.02	0.90	-0.08	0.02	0.92
ZSchool ave. poverty							-0.07	0.05	0.94
ZSchool Average ACT							0.47	0.05	1.60
Intercept	-0.515	0.08	0.60	-0.70	0.09	0.50	-0.56	0.06	0.57
Variance Components	**Variance of coefficients across high schools** *In Standard Deviations*								
ZGPA	0.110	0.246		0.096	0.204		0.112	0.364	
$ZGPA^2$	0.106	0.080		0.105	0.036	*	0.112	.037	*
ZACT	0.213	0.002	**	0.206	.000	***	0.192	.002	**
$ZACT^2$	0.089	>.500		0.088	>.500		0.066	>.500	
Intercept	0.622	0.000	***	0.859	0.000	***	0.324	0.000	***

*p<0.05, **p<0.01, ***p<0.001. Student background and college institutional control variables were grand-mean centered in all models. Variables beginning with Z were standardized, except squared terms which are the square of the standardized variables.

Figure 1. College Graduation Rates by HSGPA and ACT Score,
Controlling for Student Background and College Characteristics
Each gray line represents a high school, the black line is the average across high schools



Note: Graduation rates by school are calculated from 2-level hierarchical models that allow the relationship between ACT scores or HSGPA to vary by high school and include a quadratic term, and control for student race, ethnicity, neighborhood SES, college size, percent full-time students, student-faculty ratio, and institutional graduation rate. The average for each point reflects the predicted graduation rate given the average HSGPA or ACT score of students in a particular achievement range at each school, which is not always the midpoint. Lines only include HSGPA and test score ranges that are observed at the high school, among their college enrollees. The overall rate is calculated from a non-parametric model in which HSGPA or ACT scores are entered as a series of dummy variables, along with the same control variables.

Appendix

Table A1. College Graduation Rates by HSGPA and ACT Score
Unadjusted for Student Backgrounds, College Characteristics or High School Effects

ACT Score ->	0 to 13		14-16		16-17		18-19		20-21		22-23		24-25		26-27		28-29		30+		Overall	
HS GPA	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n	Grad	n
<1.5	11%	64	9%	115	17%	133	12%	154	16%	127	15%	55	15%	27	31%	16	N/A	3	N/A	3	14%	697
1.5-1.74	9%	57	10%	89	15%	158	14%	161	17%	104	30%	61	20%	30	13%	15	N/A	9	N/A	1	15%	685
1.75-1.99	9%	68	15%	176	21%	206	22%	247	26%	185	34%	106	37%	60	23%	26	47%	15	N/A	4	23%	1093
2.0-2.24	10%	86	21%	287	23%	379	29%	347	31%	269	43%	166	44%	97	40%	48	50%	18	36%	11	28%	1708
2.25-2.49	18%	92	21%	262	28%	453	34%	447	41%	376	56%	207	55%	150	47%	78	58%	38	44%	25	36%	2128
2.5-2.74	18%	83	34%	272	31%	477	41%	475	47%	386	57%	292	60%	218	60%	126	76%	46	67%	24	44%	2399
2.75-2.99	24%	58	33%	217	39%	429	48%	483	53%	436	63%	320	71%	275	73%	163	74%	72	73%	44	53%	2497
3.0-3.24	33%	49	40%	195	44%	392	56%	465	67%	380	77%	361	79%	282	82%	183	87%	107	78%	55	64%	2469
3.25-3.49	34%	32	45%	101	51%	273	61%	313	65%	316	73%	309	84%	210	84%	189	90%	124	84%	83	68%	1950
3.5-3.74	N/A	7	65%	43	51%	130	67%	203	73%	233	85%	202	90%	204	91%	136	92%	121	93%	103	79%	1382
3.75 and higher	N/A	5	64%	14	70%	40	71%	56	77%	129	92%	99	90%	119	94%	80	91%	70	96%	133	86%	745
Overall	17%	601	27%	1771	33%	3070	42%	3351	50%	2941	63%	2178	70%	1672	73%	1060	81%	623	83%	486	49%	17753

Graduation rates for cells with less than 10 students are not displayed to protect confidentiality.

Educational Evaluation and Policy Analysis
June 2015, Vol. 37, No. 2, pp. 206–223
DOI: 10.3102/0162373714537350

The Test-Optional Movement at America's Selective Liberal Arts Colleges: A Boon for Equity or Something Else?

Andrew S. Belasco
University of Georgia
College Transitions LLC
Kelly O. Rosinger
James C. Hearn
University of Georgia

The test-optional movement in the United States emerged largely in response to criticism of standardized admissions tests as inadequate and potentially biased measures of postsecondary promise. Although anecdotal reports suggest that test-optional policies have improved campus diversity, empirical research has not yet confirmed this claim. Consequently, this study employs quasi-experimental techniques to assess the relationship between test-optional policy implementation and subsequent growth in the proportion of low-income and minority students enrolling at adopting liberal arts colleges. It also examines whether test-optional policies increase institutional standing through greater application numbers and higher reported Scholastic Aptitude Test (SAT) scores. Results show that, on average, test-optional policies enhance the perceived selectivity, rather than the diversity, of participating institutions.

Keywords: *test-optional, admissions, longitudinal studies, administration, policy analysis*

When the first Scholastic Aptitude Test (SAT) was administered in 1926 (Gambino, 2013), advocates promoted the test as a measure of intellect and a mechanism of educational and social opportunity. At a time when access to higher education was largely determined by status, the SAT aimed to distinguish academic aptitude from "accidents" of birth and fortune and to identify talented students who would otherwise have gone unnoticed (Lemann, 1999). With the arrival of the SAT, a new meritocratic system emerged, one that promised to sort students into college on the basis of academic potential rather than social status (Jencks & Riesman, 1968; Karabel, 1984; Katz, 1978). Over the next 30 years, use of the SAT at U.S. colleges and universities increased dramatically, and by the late 1950s, the test was being administered to more than half a million high school students annually. In 2012, the number of students taking the SAT and/or American College Testing (ACT) exceeded 1.6 million in 2012, with many students taking both exams and taking the SAT and/or ACT more than once to increase scores (Lewin, 2013). Currently, most 4-year colleges and universities use standardized test scores as one factor in making admissions decisions.

Given their role in the college admissions process, standardized tests have been the subject of extensive research, and many studies have attempted to measure the predictive validity of these increasingly influential exams. Some research suggests that the SAT, coupled with high school grade point average (GPA), provides a better prediction of a student's future academic performance than high school GPA alone (Sackett et al., 2012; Shaw, Kobrin, Patterson, & Mattern, 2012). However, other studies have challenged the SAT as a reliable predictor of future college success (Crouse & Trusheim, 1988; Geiser & Studley, 2002; Rothstein, 2004), and have highlighted the persistent and positive relationship between standardized test performance and socioeconomic background as well as disparities in performance by race (Blau, Moller, & Jones, 2004; Camara & Schmidt, 1999; Fischer et al., 1996; Freedle, 2003). This latter body of research has prompted some colleges to question whether reliance on standardized testing has reinforced the exact college-related barriers that initial proponents of the SAT intended to eradicate (Epstein, 2009).

Consequently, support for the SAT, ACT, and similar standardized tests has waned at a small, but growing number of institutions, and a "test-optional movement" has emerged, particularly among liberal arts colleges, many of which have sought to eliminate or de-emphasize the use of standardized tests in the admissions process. Today, more than 50 selective liberal arts colleges have adopted test-optional admissions policies, along with approximately 800 other institutions across the United States (FairTest, 2013).

Despite public claims that test-optional policies have improved socioeconomic and racial diversity, some have questioned the motives of test-optional colleges and believe that test-optional admissions policies constitute yet another strategy to raise an institution's rank and admissions profile (Diver, 2006; Ehrenberg, 2002; Hoover, 2010). In this article, we explore both the generally stated goals of test-optional policies—expanding college opportunity and diversity—and the criticism that these policies are implemented merely to promote greater institutional standing. More specifically, we employ a difference-in-differences (DiD) analytical approach to examine whether test-optional admissions policies have achieved a commonly stated objective of increasing low-income and minority student enrollment, and also whether such policies have led to increased institutional status in the form of greater application numbers and higher reported test scores. To that end, our study addresses four research questions:

Research Question 1: Do colleges enroll significantly more (or less) low-income students (measured by Pell Grant recipient enrollment) after adopting test-optional admissions policies?

Research Question 2: Do colleges enroll significantly more (or less) underrepresented minorities after adopting test-optional admissions policies?

Research Question 3: Do colleges experience a significant rise (or decline) in freshman year applications after adopting test-optional admissions policies?

Research Question 4: Do colleges report significantly higher (or lower) average test scores after adopting test-optional admissions policies?

Literature Review

Although standardized tests assume a conspicuous role in the current college landscape, they were not widely used by postsecondary institutions until the mid-20th century, when the GI Bill of 1944 and subsequent growth in the 18- to 24-year-old population prompted an unprecedented rise in the demand for postsecondary education. Between 1950 and 1970—commonly referred to as the era of "college massification"—enrollment in U.S. higher education grew nearly fivefold (Gumport, Iannozzi, Shaman, & Zemsky, 1997). As college applications surged across the United States, selective colleges, in particular, were compelled to adopt new screening methods to sort through larger, more competitive, and increasingly heterogeneous applicant pools (Alon & Tienda, 2007; Lemann, 1999; Posselt, Jaquette, Bielby, & Bastedo, 2012); and many such institutions began to rely on standardized testing as one admissions screening mechanism.

Although the SAT and ACT originally were designed to promote college access—specifically, by identifying academically talented students, regardless of background—there has been much debate surrounding the predictive validity of these exams. Previous research has revealed a positive correlation between SAT scores and postsecondary GPA, and has also indicated that standardized test scores, in conjunction with high school GPA, serve as a better predictor of first-year academic performance than high school GPA alone (Kobrin, Patterson, Barbuti, Mattern, & Shaw, 2008; Sackett et al., 2012). However, other research contends that standardized tests have become proxies for privilege and have perpetuated class and race divisions within postsecondary education (e.g., see Grodsky, Warren, & Felts, 2008, for review of educational testing and social stratification). Several studies have cited a strong positive correlation between standardized test achievement and socioeconomic status (SES; Blau et al., 2004; Camara & Schmidt, 1999; Fischer et al., 1996; Freedle, 2003; Rothstein, 2004), and also between standardized test achievement and White racial status (Camara & Schmidt, 1999; Rothstein, 2004); while other research has suggested that standardized test scores lose much of their ability to predict postsecondary success (i.e., first-year GPA) when student SES (Geiser & Studley, 2002) and high school racial and socioeconomic diversity (Rothstein, 2004) are considered. These findings may be attributed, at least in part, to the fact that socioeconomically advantaged students are more likely to purchase test preparation materials, enroll in test preparation classes, hire a tutor, and engage in other activities that are likely to boost test scores (Buchmann, Condron, & Roscigno, 2010; Park, 2012). Finally, other critiques suggest that test scores—when compared with other measures of academic achievement, such as high school GPA or class rank—are insufficient gauges of motivation, inquisitiveness, and other qualities that contribute to learning and success (Atkinson & Geiser, 2009; Hoffman & Lowitzki, 2005).

Despite extensive research challenging the predictive validity of standardized tests, there are several recent studies indicating that the SAT and ACT continue to predict academic performance, even when background is considered (e.g., Bettinger, Evans, & Pope, 2011; Sackett, Kuncel, Arneson, & Waters, 2009; Sackett et al., 2012). For example, Sackett and colleagues (2012) found in an analysis of three large-scale datasets that the association between SAT scores and first-year academic performance decreases only slightly when socioeconomic background is considered, suggesting that the SAT remains a useful predictor of future academic achievement. In addition, Bettinger et al. (2011) discovered that ACT subscores in English and mathematics are highly predictive of first-year and second-year college GPA, even after controlling for race, gender, and (college) campus fixed effects.

While education researchers debate the merits of standardized testing, the overwhelming majority of selective colleges and universities continue to hold firm to their standardized testing requirements and use standardized test scores, among other academic and extracurricular factors, in making admissions decisions. In fact, many selective institutions have become more reliant on standardized testing in recent decades. Alon and Tienda (2007), for example, used data from two nationally representative studies to discover that, on average, America's most selective schools ascribe more weight to test scores than grades when evaluating applicants. Alon and Tienda attribute increased dependence on test scores to the perceived need for a standardized metric that is able (or that claims to be able) to identify the "aristocracy of talent" among an ever-growing pool of qualified applicants; however, they and others (Ehrenberg, 2002; Epstein, 2009) also attribute increased reliance to the rising prominence of college rankings systems, such as those released by *U.S. News & World Report*. Although contributing a relatively small percentage to the magazine's ranking formula (7.5% to 8.125% in recent years), average institutional SAT/ACT score is the largest predictor of *U.S. News* rank (Webster, 2001), and its influence may be subsumed within other measures that *U.S. News* uses to determine an institution's rank score, such as academic reputation (as reported by college administrators and high school counselors).

Indeed, enrollment managers and admissions officers face increasing pressure to enroll classes with stronger academic credentials each year. These institutional pressures have resulted in several recent cases of institutional test scores being misrepresented or deliberately manipulated for institutional purposes (e.g., Fuller, 2012; Hoover, 2012a; Supiano, 2012) Consequently, given their influence and the "elasticity of admissions data" (Hoover, 2012b), standardized test scores have been assigned considerable, and perhaps undue, emphasis in the admissions process, especially by institutions seeking to improve their standing in the rankings hierarchy.

While selective colleges, in general, have exhibited a stronger commitment to standardized testing over time; there is a growing minority of competitive institutions, primarily within the liberal arts sector, which has decided to de-emphasize or eliminate the use of standardized test scores in the admissions process. Interestingly, the test-optional "movement" among liberal arts colleges began in earnest after the speech of a university president, University of California's (UC) Richard Atkinson, who declared to the American Council on Education that overreliance on the SAT was "distorting educational priorities and practices" (Atkinson, 2001). Although UC never implemented Atkinson's recommendation that the university system abandon its SAT I admission requirement, Atkinson's speech prompted the College Board to redesign the SAT, which featured a new writing section and de-emphasized assessing student aptitude in favor of testing student preparation (Epstein, 2009). The speech also prompted scores of selective liberal arts colleges to abandon or de-emphasize standardized testing requirements in their admission processes (Epstein, 2009). Over the past decade, and despite the release of a revised SAT, more than 50 liberal arts colleges identified by Barron's *Profile of American Colleges* as "very competitive," "highly competitive," or "most competitive" have adopted test-optional policies that allow applicants to choose, without penalty, whether or not to submit their SAT or ACT scores.

In addition to expressing concerns about the biases and validity of standardized assessments, test-optional colleges commonly report that test-optional policies enhance the ethnic and economic diversity of their respective campuses without compromising the academic quality or performance of their student bodies (Bates College, 2004; Jaschik, 2006; McDermott, 2008). Espenshade and Chung's (2011) simulation study supports such claims, suggesting that test-optional policies would lead to an increase in the percentage of Black, Hispanic, and low-SES students at adopting institutions; however, it relied on predicted probabilities of admission to make assertions about *yield*, even though acceptance does not necessarily result in enrollment, especially in the case of underrepresented populations (Smith, Pender, & Howell, 2013).

To date, few studies have assessed the relationship between test-optional policies and campus diversity. Moreover, we know little about whether the implementation of test-optional policies leads to benefits that are less altruistic and more institution-specific. Several higher education leaders and reports have argued that colleges adopt test-optional policies to increase institutional status and selectivity (Ehrenberg, 2002; Epstein, 2009; Yablon, 2001), specifically through higher application numbers and reported standardized test scores. Case studies examining individual institutions' test-optional policies provide some evidence that the adoption of these policies results in increased applications from students who might otherwise not have applied (e.g., Bates and Providence colleges; Epstein, 2009). One such study of Mount Holyoke College revealed that students "underperforming" on the SAT were more likely to withhold their results from the test-optional college (Robinson & Monks, 2005), leading to higher institution-reported SAT scores. However, there have been no broad studies (i.e., studies focusing on multiple colleges) examining the effects of test-optional adoption. Thus, we know little about how the test-optional movement as a whole has influenced the admissions and enrollment profiles of participating colleges.

Conceptual Framework

To conceptualize how test-optional policies might influence admissions and enrollment at liberal arts colleges, we consider the overt and less overt intentions of test-optional adoption. To do so, we draw upon Merton's influential understanding of the manifest and latent functions of social action (e.g., Merton, 1957). Merton's approach allows us to examine the intended (manifest) and unintended (latent) functions of social policies, and how these functions serve to maintain and reinforce the current social structure and its existing inequalities (Merton, 1936, 1957).

Manifest functions refer to the intended and recognized purposes of test-optional policies. These manifest functions are institutions' commonly stated goals for adopting policies that de-emphasize or eliminate the use of test scores. Institutions that have adopted test-optional policies often cite efforts to improve diversity and to "level the playing field" for groups of students who, on average, tend to be disadvantaged by higher education's reliance on standardized testing, (Cortes, 2103; Epstein, 2009; Espenshade & Chung, 2011). By encouraging a more holistic review of applicants, test-optional admissions policies are intended to reduce the inequalities in college access that standardized test scores arguably promote. Analyzing the manifest functions of test-optional policies thus allows us to determine whether these policies have achieved a commonly stated goal of increasing postsecondary opportunity through enhancing campus economic and ethnic diversity—at liberal arts colleges specifically.

Although previous research often focuses on the recognized outcomes of test-optional policies, we extend our understanding of these policies by considering the unintended or unrecognized outcomes, or latent functions, that test-optional policies fulfill. As Merton (1957) suggested, the analysis of latent functions provides a particularly interesting area of sociological inquiry by considering how less overt outcomes enable institutions to maintain their current social position. Although test-optional admissions policies largely are hailed as efforts to expand access at selective institutions, it is also possible they serve a less noted purpose of increasing institutional status and perceived selectivity.

In a 2006 op-ed to the *New York Times*, former president of Reed College, Colin Diver, called attention to possible ulterior motives behind test-optional adoption. In his piece, Diver (2006) suggested that under test-optional policies, low-scoring students would choose not to submit their test scores, and as a consequence, test-optional colleges would increase their average institutional test scores and standing in the *U.S. News* rankings. Diver and others (e.g., Ehrenberg, 2002) also argued that institutions adopting policies that de-emphasize the use of standardized test scores encourage more applications from students who may otherwise have not applied on the basis of a test requirement or average test score.

Finally, and as Diver (2006) and Epstein (2009) noted, institutions may be aware of the implications that test-optional policies have for both enrollment and status. It is possible that college administrators may consciously adopt these policies with an eye toward increasing diversity *and* appearing more selective. If so, what may seem latent to others may actually be a manifest function and motivating factor that shapes the admissions policies administrators choose to adopt. That is, test-optional admissions policies may constitute a "double play" strategy (Bourdieu, 1996, p. 271) institutions use to promote social aims and subtly influence institutional standing. If this assessment proves accurate, test-optional policies may ultimately reaffirm the position of selective institutions, and their role in maintaining and reproducing stratification within higher education and society more broadly (Bourdieu, 1993; Bourdieu & Wacquant, 1992).

Hence, in this analysis, we examine the possibility that although test-optional policies overtly seek to expand educational opportunity, they may also result in better institutional position through increased numbers of applications and higher reported SAT/ACT scores for use in institutional rankings. Thus, in Merton's account, even if test-optional policies fail to achieve their manifest functions, institutions may still adopt or continue these policies because they fulfill a desirable latent function of increasing institutional standing.

Data and Sample

To assess how test-optional policies shape diversity and admissions profiles at liberal arts colleges, we collected time-series, cross-sectional (i.e., panel) data on 180 selective liberal arts colleges in the United States. Our panel spans nearly two decades, from 1992 to 2010, and includes annual institution-level data on several outcomes of interest, namely, the percentage of students receiving a Pell grant (any dollar amount), the percentage of students identifying as an underrepresented minority (African American, Hispanic, or Native American), the number of freshman applications submitted to an institution, and an institution's average reported SAT score (25th percentile, critical reading, and math combined). Our primary independent variable is dichotomous and indicates whether colleges in the sample possess a test-optional admissions policy during a given year. We assign test-optional status only to those colleges that have made the submission of *all* test scores optional for *all* students, *and* that do not penalize applicants who wish to withhold their test scores. For example, several liberal arts colleges have adopted test-flexible admissions policies—that do not require SAT scores, but that still require applicant scores from one or several other standardized tests (e.g., ACT, Advanced Placement [AP], or SAT subject tests)—and/or have made the submission of test scores optional for only a small subset of high-achieving students. These colleges cannot be considered test-optional in a definitional sense and are designated as "test-requiring" for the purposes of this study.

In addition to our dependent and primary independent variables, we also include controls for several time-variant variables that are likely to influence the diversity and admission profile of a liberal arts college, specifically full-time enrollment (FTE), annual tuition and fees, institutional grant award per FTE, education and related expenditures per FTE, admission rate, and a dichotomous variable indicating whether an institution adopted a no-loan financial aid policy in a given year. Financial measures are adjusted for inflation using the Consumer Price Index to reflect 2010 dollars and are logged to ease interpretation and provide a more normal distribution to the data.

Data incorporated into the panel come from multiple postsecondary data sources, including the U.S. Department of Education, the Integrated Postsecondary Education Data System (IPEDS), the Delta Cost Project, and the College Board's (2011) *Annual Survey of Colleges*. The data encompass years before and after test-optional "treatment," thereby providing a suitable data space within which to employ DiD modeling.

A quasi-experimental technique, DiD, employs a fixed-effects strategy to isolate group- or aggregate-level changes resulting from a particular intervention or policy. Specifically, DiD exploits time-induced variation to control for potential observed and unobserved differences that exist across treated and control groups and which may obscure effects that are attributed to the treatment itself (Gelman & Hill, 2006). In this study, DiD allows us to assess whether test-optional colleges experienced significant changes in the above-mentioned outcomes after adoption of their respective policies, controlling for potentially confounding time trends and pre-existing differences between test-optional and test-requiring institutions.

To reduce bias and meet identifying assumptions of the DiD model (discussed further below), we limit our sample to liberal arts colleges that *Barron's* Admissions Competitive Index categorizes as "competitive," "very competitive," "highly competitive," or "most competitive." Institutions at which standardized tests are not likely to figure prominently in the admissions process are excluded from the analysis, specifically institutions that are classified by *Barron's* as "less competitive," "non-competitive," or "special"—all of which have relatively high acceptance rates (more than 85%), admit applicants with low standardized test scores, and/or admit applicants largely on the basis of non-academic credentials. In addition, we focus our analysis on liberal arts colleges, in particular, because, during the period of our study, test-optional policies were adopted primarily by institutions in this sector.[1] Table 1 lists the test-optional liberal arts colleges within our panel and the academic year (ending) in which test-optional policies were adopted.

Analytic Technique

In cross-sectional evaluations of test-optional initiatives, estimated effects may confound policy-related gains in diversity and admissions profile with unobservable, institution-level attributes, which may also contribute to these outcomes, such as a college's culture or academic environment. Likewise, a pure time-series analysis may uncover a significant post-policy effect, but the effect may be spurious due to time trends that move most or all colleges to experience a change in their Pell rates or reported SAT scores, for example. In contrast, DiD controls for enrollment trends *and* pre-treatment differences between institutions, in effect, using both as baselines against which to compare the after-intervention outcomes of test-optional and test-requiring schools. This enables us to distinguish whether, and to what extent, post-implementation effects are attributable to the test-optional policy itself. The DiD model is formally expressed as

$$Y_{cy} = \beta_0 + \beta_1 T_c + \beta_2 A_{cy} + \gamma \mathbf{X}_{cy} + \delta_1 T_c A_{cy} + \varepsilon_{cy}, \quad (1)$$

where Y_{cy} is an outcome of interest; T_c is a dichotomous measure indicating whether a college, c, received the test-optional "treatment" during any year in the panel, y, and captures pre-treatment differences between optional and non-optional schools; A_{cy} is a dichotomous measure

TABLE 1
Sample Liberal Arts Colleges Adopting Test-Optional Policies

College (City, State)	Year of Adoption (Ending)
Wheaton College (Wheaton, MA)	1993
Dickinson College (Carlisle, PA)	1995
Hartwick College (Oneonta, NY)	1996
Muhlenberg College (Allentown, PA)	1997
Mount Holyoke College (South Hadley, MA)	2002
Pitzer College (Claremont, CA)	2004
Sarah Lawrence College (Bronxville, NY)	2005
Chatham University (Pittsburgh, PA)	2006
College of the Holy Cross (Worcester, MA)	2006
Knox College (Galesburg, IL)	2006
Lawrence University (Appleton, WI)	2006
St. Lawrence University (Catnon, NY)	2006
Susquehanna University (Selinsgrove, PA)	2006
Bennington College (Bennington, VT)	2007
Drew University (Madison, NJ)	2007
Eckerd College (St. Petersburg, FL)	2007
Franklin & Marshall College (Lancaster, PA)	2007
Gettysburg College (Gettysburg, PA)	2007
Guilford College (Greensboro, NC)	2007
Gustavus Adolphus College (St. Peter, MN)	2007
Hobart and William Smith Colleges (Geneva, NY)	2007
Juniata College (Huntingdon, PA)	2007
Lake Forest College (Lake Forest, IL)	2007
Lycoming College (Williamsport, PA)	2007
Union College (Schenectady, NY)	2007
Augustana College (Rock Island, IL)	2008
Denison University (Granville, OH)	2008
Wittenberg University (Springfield, OH)	2008
Albright College (Reading, PA)	2009
Goucher College (Towson, MD)	2009
Marlboro College (Marlboro, VT)	2009
Smith College (Northampton, MA)	2009

equaling "1" in years during and after implementation of a test-optional policy and captures changes in our outcomes of interest that may have occurred in the absence of a test-optional policy; $\mathbf{X}_{cy}$ indicates a vector of relevant covariates described above; and δ_1, the coefficient of interest, interacts with the intervention and time indicators and represents the DiD estimate, where

$$\delta_1 = (Y_{Treat(after)} - Y_{Treat(before)}) - (Y_{Control(after)} - Y_{Control(before)}), \qquad (2)$$

which represents the difference in outcomes between the pre- and post-policy time periods, while controlling for pre-existing differences in outcomes between test-optional and test-requiring institutions.

Given the standard ordinary least squares (OLS) formulation of the above model, it is necessary to account for characteristics of our data and sample, which could lead to bias and/or inefficient estimates, even within the DiD framework. First, given that colleges instituted test-optional policies in different years, the simplified model in Equation 1 may over- or underestimate the effect of test-optional intervention as it assigns treatment to colleges that did not yet implement a test-optional policy. As a corrective measure, we incorporate institution- and year-fixed effects to specify the exact year in which a participating school received intervention and, in contrast to the simplified model in Equation 1, to account for variation in the duration of "treatment" among test-optional colleges (Bertrand, Duflo, & Mullainathan, 2004; Dynarksi, 2004). In particular, we estimate the following revised model, which should provide more refined evidence of test-optional effects:

$$Y_{cy} = \alpha A_c + \beta B_y + \gamma \mathbf{X}_{cy} + \delta_1 T_{cy} + \varepsilon_{cy}, \qquad (3)$$

where A_c and B_y are fixed effects for colleges, *c*, and years, *y*, respectively; $\mathbf{X}_{cy}$ represents a vector of included covariates; ε_{cy} is an idiosyncratic error term; and δ_1 is our coefficient of interest and equal to "1" in any academic year when an institution's incoming class of students benefitted from a test-optional admission policy. For example, if a college adopted a test-optional admissions policy during the 2004–2005 academic year for the incoming class of 2005–2006, the institution is first indicated as a test-optional college in the 2005–2006 academic year, as 2005–2006 is the first year in which test-optional policies may affect institutional indicators, such as Pell rates, minority rates, average test scores, and *reported* application numbers.[2]

In addition, given that our analysis encompasses multiple years before and after test-optional "intervention," we also conduct a series of Durbin–Watson tests, which yield evidence of serial correlation in the simple and revised models (Equations 1 and 3, respectively) for all outcomes. To correct for possible Type 1 error, we incorporate cluster-robust standard errors into each of our models (White, 1980), which adjust the estimated variance–covariance matrix to account for correlated residuals within clusters (i.e., colleges) and which should provide for efficient estimates of a test-optional effect, especially given that our sample has a *N* greater than 50 (Bertrand et al., 2004).

Finally, after estimating both models, we explore whether our DiD design meets the assumption of parallel trends. To yield unbiased estimates, DiD models must meet the strong assumption that treated and control groups would exhibit parallel trends in the absence of intervention (Angrist & Pischke, 2009)—which, according to Abadie (2005), "may be implausible if pre-treatment characteristics that are thought to be associated with the dynamics of an outcome variable are unbalanced between the treated and untreated group" (p. 2).

Potentially, there are differences between test-optional and test-requiring colleges not accounted for by Equation 3, and which may influence selection into "treatment," as well as the direction and rate at which outcomes among the two groups change. While pre-intervention data and the aforementioned covariates control for at least some of these differences, there may be other influential variables omitted from our models, which could potentially preclude accurate estimation of a test-optional effect.

Causal inference via DiD requires that we construct an appropriate counterfactual scenario where treated units (i.e., test-optional colleges) are instead assigned to the control group (i.e., test-requiring colleges), and vice versa—because any unit can be observed under only one of two conditions. To infer a causal effect of test-optional intervention, we must adequately approximate the outcomes of a "treated" college under control conditions (i.e., if it did not participate in test-optional admissions). If we can construct this counterfactual condition or "what if" scenario for treated units in our sample, we can *estimate* the average treatment effect of the test-optional policy: $E[Y_{1c} - Y_{0c}]$. Doing so, however, requires that we compare test-optional schools with "control" schools, which, given their characteristics and context, would exhibit similar trends in the absence of test-optional "treatment." If treated and control colleges within our sample differ on particular unobservables that lead to diverging outcomes, regardless of intervention, we cannot determine whether or which portion of a potential test-optional effect is attributable to the policy itself or to another difference, policy change, or event that is not accounted for by our model and that may also influence selection into treatment or our outcomes.

Although the parallel trends assumption is not formally testable, we adopt three techniques to examine whether parallel trends criteria have been met. First, and as indicated previously, we estimate each model on a disaggregated sample of colleges that share similar institutional characteristics and that are most likely to adopt test-optional policies, namely selective, liberal arts colleges. Restricting our sample to institutions of the same sector and similar selectivity levels should provide sufficient overlap (i.e., a range of common support) between test-optional and test-requiring schools, and consequently, allow us to extrapolate counterfactual outcomes via a DiD regression.

Second, we add an institution-specific trend to our set of covariates (Angrist & Pischke, 2009), which controls for the possibility that test-optional and test-requiring schools may have experienced different admissions- and campus-related trends prior to policy implementation. Trend variables are created by regressing each dependent variable on year, for each institution, using data from 1992 to 1995, the period before all but one institution in our dataset adopted a test-optional policy.[3] The trend variables incorporated into our models multiply the resulting coefficients by year and are unique for each institution-year, and as such, allow institutions to follow a different trend throughout the panel. If estimated effects are robust, the inclusion of institution-specific trends should not alter the magnitude or significance of the coefficients of our test-optional indicator.

Finally, after estimating our models, we conducted a series of placebo tests to confirm that effects are evident only after policy implementation and are not the result of some other factor unaccounted for by Equation 3 (Bertrand et al., 2004). To carry out placebo testing, we estimate models for each outcome, including only panel data for years before test-optional intervention (1992–1995), and then assign test-optional "treatment" to colleges in all years after 1992. We anticipate that placebo models indicating treatment in 1993, 1994, and 1995 will yield insignificant effects of a test-optional policy, because policy implementation is synthetic and never actually occurs. However, if our test-optional indicator is significant, we must consider that effects attributed to the outcome being modeled are spurious (and possibly null), and that changes in the outcome, if any, are due to other unobservable measures.

Limitations

Despite the application of several bias-reducing techniques, this study is still limited in three important ways. First, there are several colleges for which we were unable to collect pre-adoption data. Five colleges, namely Bard, Bates, Bowdoin, Hampshire, and Lewis and Clark, implemented test-optional policies before 1992 and as early as 1965. While efforts were made to collect data prior to 1992, inconsistencies in IPEDS reporting (for grant awards and minority enrollment) and missing College Board data (for SAT scores and freshman applications) prevented us from expanding our panel to earlier years. Although "early-adopting" colleges constitute a small percentage of all test-optional colleges, and adopted policies prior to, and irrespective of, the test-optional movement, their influence could shed light on the long-term influence of test-optional initiatives. With this in mind, additional research might explore other techniques to examine test-optional-related changes among this unique group of institutions.

Second, while our fixed-effects identification strategy controlled for time-invariant omitted variables that may confound the institution-related effects of test-optional policies, it did not control for variables that change over time, which were not incorporated into our models and which may ultimately confound our estimates. For example, given the inconsistencies in endowment reporting during the period of our study, we were unable to include a variable for each college's annual institutional endowment—a potentially important indicator of campus diversity and admissions competitiveness. Although we collected data on an adequate proxy, institutional grant award per student, there may still be other elements of endowment that contributed to our outcomes of interest, above and beyond what is used for financial aid. In addition, a measure indicating the percentage of students submitting test scores may have provided for finer distinctions between test-optional programs and a more nuanced discussion on the relationship between test-optional "participation" and our dependent variables; however, reliable data for this indicator were not available.

Finally, several variables have missing data, specifically those for Pell rate (0.85%), reported SAT score (1.81%), applications (2.31%), and acceptance rate (2.31%). As a robustness check, we imputed missing values using chained equations and compared the results of our models with imputed data against our original models (with missing data). Our results remained the same; however, our findings may still be susceptible to non-response bias, especially because the majority of missingness occurs within a particular time frame, namely the first 5 years of our panel.

Results

The graphs in Figure 1 illustrate changes in institutional diversity and admissions profile during the period of our study for both test-optional and test-requiring colleges. Graphs A and B show, respectively, that test-optional colleges enrolled a lower proportion of Pell recipients and underrepresented minorities, on average, than test-requiring institutions—during all years of the panel. Furthermore, and somewhat to our surprise, Graphs A and B reveal that test-optional colleges did not make any progress in narrowing these diversity-related gaps after they adopted test-optional policies. In contrast, Graphs C and D suggest that test-optional adopters did achieve relative gains on certain admissions-related indicators. For example, while test-optional institutions reported higher average SAT scores in initial years of the panel, their margins increased in later years, by approximately 25 points on average, as Graph C shows.[4] Graph D also depicts steadily increasing margins in application totals between test-optional and test-requiring

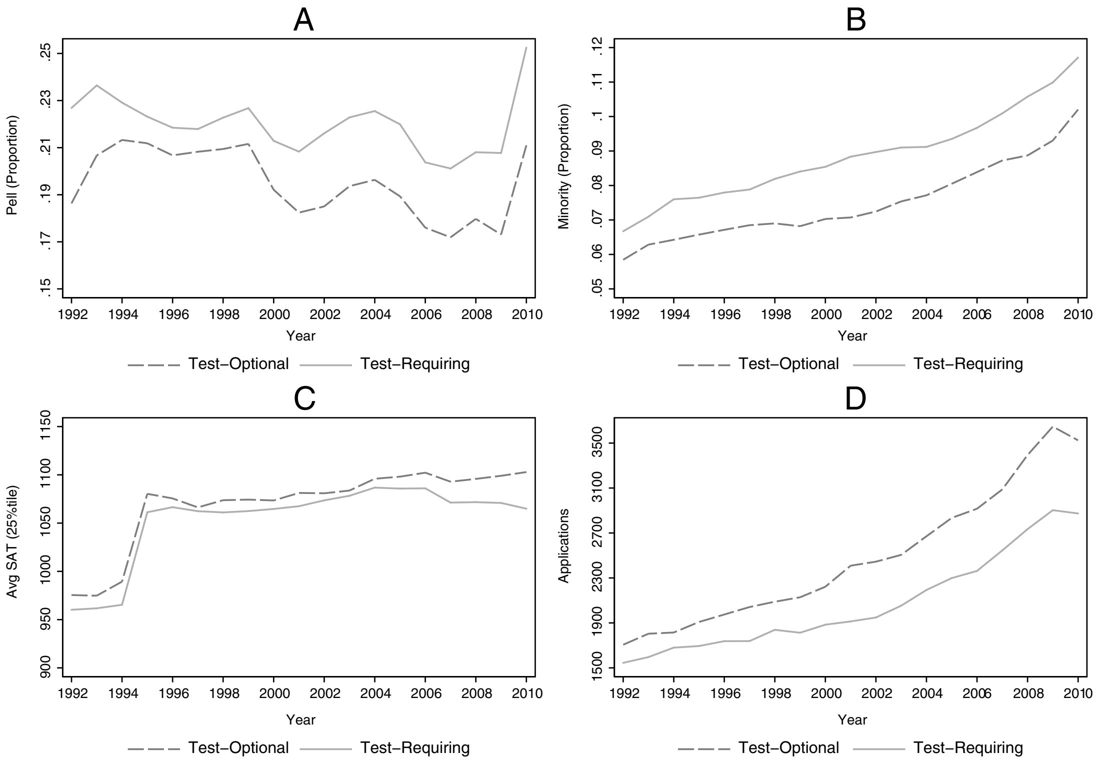


FIGURE 1. *Institutional diversity and admissions profile: Averages for test-optional and test-requiring colleges (1992–2010).*
Note. SAT = Scholastic Aptitude Test.

schools. In the first year of our panel, (eventual), test-optional colleges received 150 more applications, on average, than their test-requiring counterparts; by the end of our panel, test-optional colleges were receiving approximately 550 more applications.[5]

While the graphs in Figure 1 illuminate changes in our outcomes of interest, they cannot communicate the magnitude and significance of such changes, especially given that additional factors, besides test-optional policy implementation, may have contributed to differences in diversity and admissions-related trends between test-optional and test-requiring institutions. Indeed, the descriptive statistics in Table 2 reveal substantial growth in other institution-level indicators, which may have contributed to diverging outcomes between the two groups. For example, Table 2 shows that institutional grant dollars per FTE at test-optional colleges more than doubled in constant dollars over the course of our panel, and averaged more than US$13,000 per student by 2010, which may explain relative gains in the number of applications received at these schools. In addition, test-optional colleges experienced greater increases in tuition and fee prices in constant dollars during the period of our study, which may have prevented optimal numbers of low-income and/or minority students from applying, and consequently, may have suppressed the positive effects that test-optional policies might have otherwise had on the diversity of adopting institutions. If tuition remained constant, would test-optional policies have contributed to increases in low-income and minority enrollment—as many test-optional colleges have claimed, and despite what the graph in Figure 1 indicates? Can diverging application totals be attributed to test-optional polices, increased grant aid, or both? Results from our DiD models address these and other such questions.

Table 3 displays our regression results, which appear to confirm what the graphs in Figure 1 suggest—that test-optional admissions policies

TABLE 2
Means (and Standard Deviations) of Independent Variables (Test-Optional vs. Test-Requiring Colleges)

Variable	Minimum	Maximum	Test-optional (1992)	Test-optional (2010)	Test-requiring (1992)	Test-requiring (2010)
Independent						
No-loan policy	0.00	1.00	0.00	0.03	0.00	0.11
Undergraduate enrollment (FTE)	59.61	7,686.76	1,541.36	1,951.56	1,489.35	1,750.80
			(640.57)	(607.10)	(869.12)	(1,059.47)
E&R expenditures (per FTE)	6,744.15	97,196.20	22,861.79	29,151.73	19,753.82	27,946.33
			(5,226.24)	(7,712.33)	(6,870.28)	(11,922.24)
Tuition & fees	3,124.96	45,895.54	22,682.09	35,477.97	17,397.40	28,909.37
			(3,226.14)	(4,008.84)	(5,361.74)	(7,604.14)
Institutional grant award (per FTE)	3.26	21,933.67	6,308.39	13,358.18	4,592.59	11,494.75
			(1,667.48)	(3,079.46)	(2,214.47)	(4,588.02)
Admission rate	0.15	1.00	0.71	0.59	0.72	0.60
			(0.10)	(0.15)	(0.17)	(0.20)
Dependent						
Proportion Pell	0.03	0.82	0.19	0.21	0.23	0.25
			(0.08)	(0.08)	(0.12)	(0.12)
Proportion minority	0.00	0.56	0.06	0.10	0.07	0.12
			(0.03)	(0.05)	(0.05)	(0.07)
Applications	23	10,068	1,706.16	3,524.38	1,544.91	2,980.06
			(927.05)	(1,545.08)	(1,215.49)	(2,121.63)
Reported SAT score (25th percentile)	600	1,440	975.48	1,102.90	960.22	1,062.25
			(73.30)	(97.44)	(129.75)	(142.60)
Institutions (*N*)			32	32	148	148

Note. FTE = full-time enrollment; SAT = Scholastic Aptitude Test.

do not increase the diversity of policy-adopting liberal arts colleges, on average. In particular, when controlling for unobserved heterogeneity (via institution- and year-fixed effects) and other time-varying characteristics, test-optional policies failed to effect a positive change in the proportion of low-income and minority students enrolling at test-optional institutions. This finding contradicts simulated analyses of test-optional programs (Espenshade & Chung, 2011) and is also counter to the reports of several test-optional colleges (Bates College, 2004; Jaschik, 2006; McDermott, 2008). Yet, given the descriptive nature and narrow focus of these past studies—previous reports consisted mostly of case studies focusing on one or a small number of institutions—and the quasi-experimental nature of our own study, we are confident that results yielded from our models are robust and provide some evidence that test-optional policies overall have not been the catalysts of diversity that many have claimed them to be.

Despite their seemingly non-significant impact on racial and economic diversity, test-optional policies appear to benefit adopting colleges in other, more institution-promoting ways. As indicated in the third set of columns in Table 3, implementing a test-optional admissions policy appears to exert a positive and significant influence on the number of applications a college receives. Specifically, after controlling for fixed effects, institution-specific trends, and other influential covariates, our results suggest that liberal arts colleges receive approximately 220 more applications, on average, after adopting a test-optional policy. This constitutes a substantial increase, especially given that colleges in our sample enroll only 400 first-year students annually, on average; however, the statistical significance of our finding may have

TABLE 3

Estimating the Effects of Test-Optional Policies

Outcome	Proportion Pell			Proportion minority			Applications[a]			Reported SAT score		
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Test-optional policy	−0.009	−0.006	−0.012	−0.003	−0.004	−0.003	300.643*	234.023*	221.331*	25.664**	27.184***	25.674**
	(0.010)	(0.009)	(0.009)	(0.004)	(0.005)	(0.005)	(134.250)	(112.853)	(107.781)	(7.903)	(7.974)	(7.792)
No-loan policy		0.008	0.004		0.008	0.009		563.741***	610.153***		21.418***	23.257***
		(0.007)	(0.008)		(0.005)	(0.005)		(157.404)	(166.925)		(4.780)	(5.129)
Undergraduate FTE (ln)		−0.041	−0.041		0.026	0.026		789.916*	1,118.951**		34.410	35.541
		(0.041)	(0.043)		(0.019)	(0.017)		(405.044)	(364.396)		(20.379)	(21.243)
E&R expenditures (ln)		−0.025	−0.026		0.000	0.002		541.808*	693.587**		57.086**	59.039**
		(0.030)	(0.031)		(0.015)	(0.014)		(265.210)	(251.308)		(17.671)	(18.345)
Tuition & fees (ln)		0.002	0.008		−0.008	−0.007		291.776	220.377		22.082	15.015
		(0.020)	(0.018)		(0.008)	(0.007)		(201.747)	(179.190)		(18.054)	(18.426)
Grant/FTE (ln)		0.010	0.011		0.006	0.006		−37.337	−50.833		2.646	2.029
		(0.006)	(0.006)		(0.003)	(0.003)		(52.192)	(55.039)		(3.933)	(3.992)
Admission rate		0.020	0.026		0.003	−0.003		−2,611.427***	−2,578.454***		−63.434***	−69.176***
		(0.019)	(0.019)		(0.011)	(0.011)		(334.937)	(336.692)		(17.688)	(17.612)
Observations	3,389	3,269	3,262	3,418	3,292	3,292	3,339	3,292	3,269	3,356	3,243	3,182
R^2	.893	.891	.898	.846	.851	.854	.908	.928	.932	.917	.920	.922
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Institution-specific trend	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Placebo effect[b]	—	—	—	—	—	—	No	No	No	No	No	No

Note. Robust standard errors clustered at the institution level are reported in parentheses. SAT = Scholastic Aptitude Test; FTE = full-time enrollment.

[a]Models incorporating the square-root transformation of applications produce positive yet insignificant results.

[b]We test for placebo effects in models where the test-optional coefficient is significant.

*$p < .05$. **$p < .01$. ***$p < .001$.

more to do with our data than our test-optional indicator. Indeed, normality tests (Jarque & Bera, 1987; Royston, 1991) offered some evidence that our variable for applications was positively skewed. To partially correct for non-normality, we re-estimated our model using the square-root transformation of our "applications" measure, and found that effects for test-optional adoption were still positive but no longer significant.[6] As such, our analysis provides interesting, yet inconclusive, results on the relationship between test-optional policies and application numbers.

Finally, test-optional policies also appear to be associated with an increase in reported test scores. Consistent with the claims of past reports (Ehrenberg, 2002; Yablon, 2001), liberal arts colleges that implement test-optional policies experience a subsequent rise in their reported SAT scores, by approximately 26 points, on average, all else equal. Furthermore, the magnitude and significance of these test-related effects remain consistent across models, even after controlling for trends, other potential confounders, and possible placebo effects—suggesting that results with respect to this outcome are quite robust. In sum, findings from our analyses indicate that test-optional policies enhance the appearance of selectivity, rather than the diversity, of adopting institutions.

Discussion

Our findings suggest that test-optional admissions policies, as a whole, have done little to meet their manifest goals of expanding educational opportunity for low-income and minority students. However, we find evidence that test-optional policies fulfill a latent function of increasing the perceived selectivity and status of these institutions. In doing so, these policies may serve to reproduce and maintain the current social structure—and its inequalities—within U.S. higher education.

While this study provides evidence of how test-optional admissions policies shape diversity and admissions profiles, more broadly, it serves as a reminder of the values that are reflected in the process of selecting students into liberal arts colleges.

The SAT and other standardized tests were initially adopted to sort students according to academic ability rather than status and background. This sorting mechanism, however, favored wealthy students and reinforced their disproportionate presence at the nation's most selective institutions. In a way, the SAT became an adaptive mechanism that upper-class families used to secure their future social status (Alon, 2009)—which, in part, may explain why the SAT continues to predominate the selective college admissions process. While selective institutions have become increasingly open to considering SAT alternatives, other standardized assessments—including the ACT, Advanced Placement, International Baccalaureate (IB), and SAT subject tests—are vulnerable to the same inequities. For example, affluent students and families can often "buy" their way to improved scores on any standardized test by hiring a private tutor, enrolling in a test preparation course, and/or registering for several administrations of the same exam (Lemann, 1999; Lewin, 2013; Vigdor & Clotfelter, 2003). Previous research shows that one or more of these costly strategies usually results in improved standardized test scores and better admissions prospects at selective colleges and universities (Buchmann et al., 2010).

Despite the clear relationship between privilege and standardized test performance, the adoption of test-optional admissions policies does not seem an adequate solution to providing educational opportunity for low-income and minority students. In fact, test-optional admission policies may perpetuate stratification within the postsecondary sector, in particular, by assigning greater importance to credentials that are more accessible to advantaged populations. Without access to standardized test data for every applicant, test-optional colleges rely more heavily on school-specific measures, such as strength of curriculum or involvement outside the classroom, to draw comparisons between prospective students; however, several studies reveal that the availability of advanced (AP, IB, and honors) courses and extracurricular opportunities is unequally distributed across socioeconomic groups (Espenshade & Radford, 2009; Iatarola, Conger, & Long, 2011; Klugman, 2013; Perna et al., 2013), and that low-SES students face greater obstacles to participating in the classes and activities that facilitate selective college enrollment (Klugman, 2012). As a result, test-optional colleges may be inadvertently trading one inequitable policy for another—a troubling notion given that 11 additional selective liberal arts colleges have adopted test-optional polices in the past 2 years alone,[7] advancing what Diver (2006) referred to as a "new front in the admissions arms race."

Although implications for policy and practice are not entirely clear, our study reveals that eliminating or de-emphasizing standardized tests in the admissions process has not reduced educational inequalities, on average. These results indicate that the connection between social status and college admission is deeply embedded (Thacker, 2005), and perhaps more than the test-optional movement could have predicted. Our study also indicates that selective institutions cannot be relied upon, at least solely, to stem disparities in postsecondary access, which is not entirely surprising, given that most selective colleges and universities rely on a host of external resource providers that place significant emphasis on institutional position and rank (e.g., students, families, government, industry, etc.; Bastedo & Bowman, 2011; Meredith, 2004).

Nevertheless, if test-optional and other selective colleges are sincere in their desires to increase access and enroll more underrepresented students, they might consider acknowledging the SAT and other similar tests as imperfect yet useful indicators of academic achievement, as Diver (2006) and Epstein (2009) suggested, while learning to more appropriately situate a student's test score within his or her particular context.

Test-optional and other selective institutions might also consider reexamining their recruitment strategies. A wave of recent research on postsecondary "undermatch" reveals that a majority of high-achieving, low-income students fail even to apply at selective colleges and are generally unaware of the admissions requirements and benefits associated with selective higher education (Belasco & Trivette, in press; Hoxby & Avery, 2012; Smith et al., 2013). These findings are likely related to current recruitment practices at many selective colleges, which pay inadequate attention to the places where underrepresented students live and learn, largely ignoring geographically remote areas and/or low-income schools in favor of more cost-effective or "fruitful" locales (Hill & Winston, 2010; Stevens, 2007). Arguably, institutions that fail to reach a majority of underrepresented students, through recruitment or other outreach initiatives, will find it difficult to improve diversity in meaningful and significant ways, regardless of their admissions criteria. If test-optional and other selective colleges genuinely aim to become more inclusive, they must meet underrepresented students where they actually are, instead of where they "should be."

However, as intimated previously, achieving a more equitable approach to student recruitment and applicant evaluation will likely depend on the extent to which selective colleges can meet their market-related needs. To that end, it is important that selective institutions collaborate with other stakeholders to devise *and promote* new measures of excellence within higher education that could include the extent to which institutions enroll and graduate underrepresented students, the amount of resources institutions allocate to public service, average student debt load, and other indicators of postsecondary outcomes that demonstrate what colleges do, rather than whom they accept. Until U.S. higher education learns to distinguish excellence from prestige, institutions across all sectors will remain prone to prioritizing status over equity—merely to survive, at least.

Finally, it is important that selective institutions be more transparent and forthcoming about the extent to which they can accommodate disadvantaged populations. Most undermatch studies examining the lack of high-achieving, low-income students at selective institutions fail to discuss how selective colleges would respond to an influx of low-income applicants, for example. In this scenario, would Amherst or Pomona adjust its enrollment strategy to accommodate a significantly greater number of financially needy students? Or, is it more likely that a greater number of needy students would be competing for (roughly) the same number of seats? How would a similar scenario play out at Dickinson or Denison? Although answers to these questions may prompt contempt among the general public or lead to politically unpopular proposals—such as those recommending significant increases to federal and/or state aid for low-income students—they would propel discussion on what is really required to improve diversity at America's most competitive colleges, compelling all parties to deal in reality rather than ideals.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. A review of the Fairtest newsletter archives (www.fairtest.org) and various college websites revealed that 37 of 44 competitive institutions (as defined by Barron's) adopting test-optional policies before 2010 were liberal arts colleges.

2. The College Board commonly reports an institution's application numbers for the prior academic year. For example, application data in College Board's *Annual Survey of Colleges* labeled 2010 indicate the number of applications submitted in 2009.

3. Trend indicators for Wheaton College (Massachusetts), which adopted a test-optional admissions policy in 1993 (academic year ending), were created using data from 1992 and 1993 only, the 2 years before the institution could have experienced any "test-optional effects."

4. All colleges experienced sharp increases in their reported Scholastic Aptitude Test (SAT) scores after the College Board re-centered score scales in 1995 to provide easier performance comparisons among the contemporary test-taking population.

5. Growth in Pell rates and declines in application totals after 2009 are likely attributed to the Great Recession, and its negative influence on demand for liberal arts education.

6. Normality tests, along with descriptive statistics and histograms, show that a square-root transformation performs better than a log-transformation in allowing for more normal distribution. However, skewness and kurtosis tests still detect some non-normality within our transformed variable.

7. Including Agnes Scott College, Connecticut College, Earlham College, Furman College, Illinois College, Manhattanville College, Moravian College, St. Leo College, University of the South, Ursinus College, and Washington and Jefferson.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, *72*, 1–19.

Alon, S. (2009). The evolution of class inequality in higher education: Competition, exclusion, and adaptation. *American Sociological Review*, *74*, 731–755.

Alon, S., & Tienda, M. (2007). Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, *72*, 487–511.

Angrist, J., & Pischke, J. (2009). *Mostly harmless econometrics*. Princeton, NJ: Princeton University Press.

Atkinson, R. C. (2001, February 18). *Standardized tests and access to American universities*. The 2001 Robert H. Atwell Distinguished Lecture, delivered at the 83rd Annual Meeting of the American Council on Education, Washington, DC.

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, *38*, 665–676.

Bastedo, M. N., & Bowman, N. A. (2011). College rankings as an interorganizational dependency: Establishing the foundation for strategic and institutional accounts. *Research in Higher Education*, *52*, 3–23.

Bates College. (2004, October 10). *PowerPoint analysis: SAT submitters and non-submitters*. Retrieved from http://www.bates.edu/news/2004/10/10/powerpoint-analysis/

Belasco, A. S., & Trivette, M. J. (in press). Aiming low: Estimating the scope and predictors of postsecondary undermatch. *Journal of Higher Education*.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*, 249–275.

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2011). *Improving college performance and retention the easy way: Unpacking the ACT exam* (No. w17119). Cambridge, MA: National Bureau of Economic Research.

Blau, J. R., Moller, S., & Jones, L. V. (2004). Why test? Talent loss and enrollment loss. *Social Science Research*, *33*, 409–434.

Bourdieu, P. (1993). *The field of cultural production*. Cambridge, UK: Polity Press.

Bourdieu, P. (1996). *The state nobility: Elite schools in the field of power*. Stanford, CA: Stanford University Press.

Bourdieu, P., & Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. Cambridge, UK: Polity Press.

Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces*, *89*, 435–461.

Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification* (College Board Report No. 99-5). New York, NY: College Entrance Examination Board.

College Board. (2011). *Annual survey of colleges 1990-2010*. New York, NY: Author.

Cortes, C. M. (2013). Profile in action: Linking admission and retention. *New Directions for Higher Education*, *2013*(161), 59–69.

Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. Chicago, IL: University of Chicago Press.

Diver, C. S. (2006, September 18). Skip the test, betray the cause. *The New York Times*. Retrieved from http://www.nytimes.com/2006/09/18/opinion/18diver.html?_r=0

Dynarksi, S. (2004). The new merit aid. In C. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 63–100). Chicago, IL: University of Chicago Press.

Ehrenberg, R. (2002). Reaching for the brass ring: The U.S. News & World Report rankings and competition. *The Review of Higher Education*, *26*, 145–162.

Epstein, J. (2009, Summer). Behind the SAT optional movement: Context and controversy. *Journal of College Admission*, 9–19.

Espenshade, T. J., & Chung, C. Y. (2011). Diversity outcomes of test-optional policies. In J. A. Soares (Ed.), *SAT wars: The case for test-optional admissions* (pp. 177–200). New York, NY: Teachers College Press.

Espenshade, T. J., & Radford, A. W. (2009). *No longer separate not yet equal: Race and class in elite college admission and campus life*. Princeton, NJ: Princeton University Press.

FairTest: The National Center for Fair and Open Testing. (2013). *Test score optional list*. Retrieved from http://www.fairtest.org/university/optional

Fischer, C. S., Hout, M., Jankowski, M. S., Lucas, S. R., Swidler, A., & Voss, K. (1996). *Inequality by design*. Princeton, NJ: Princeton University Press.

Freedle, R. O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, *73*, 1–43.

Fuller, A. (2012, September 12). Median SAT scores can get murky. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/headcount/median-sat-scores-they-can-get-murky/31596

Gambino, M. (2013, April 12). Document deep dive: What was on the first SAT? *Smithsonian Magazine*. Retrieved from http://www.smithsonianmag.com/ist/?next=/history/document-deep-dive-what-was-on-the-first-sat-21720496/

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*, 1–26.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Grodsky, E., Warren, J. R., & Felts, E. (2008). Testing and social stratification in American education. *Annual Review of Sociology*, *34*, 385–404.

Gumport, P. J., Iannozzi, M., Shaman, S., & Zemsky, R. (1997). *The United States country report: Trends in higher education from massification to post-massification*. Stanford, CA: National Center for Postsecondary Improvement.

Hill, C. B., & Winston, G. C. (2010). Low-income students and highly selective private colleges: Geography, searching, and recruiting. *Economics of Education Review*, *29*, 495–503.

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review of Higher Education*, *28*, 455–474.

Hoover, E. (2010, March 21). Colleges explore shades of gray in making entrance tests optional. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Colleges-Explore-Shades-of/64758/

Hoover, E. (2012a, January 31). Claremont McKenna official resigns after falsely reporting SAT scores. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/headcount/claremont-mckenna-official-resigns-after-falsely-reporting-sat-scores/29556

Hoover, E. (2012b, February 1). Inflated SAT scores reveal "elasticity of admissions data." *Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/headcount/inflated-sat-scores-reveal-elasticity-of-admissions-data/29575

Hoxby, C. M., & Avery, C. (2012). *The missing "One-Offs": The hidden supply of high-achieving, low income students* (No. w18586). Cambridge, MA: National Bureau of Economic Research.

Iatarola, P., Conger, D., & Long, M. C. (2011). Determinants of high schools' advanced course offerings. *Educational Evaluation and Policy Analysis*, *33*, 340–359.

Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, *55*, 163–172.

Jaschik, S. (2006, May 26). Momentum for going SAT optional. *Insider Higher Ed*. Retrieved from http://www.insidehighered.com/news/2006/05/26/sat

Jencks, C., & Riesman, D. (1968). *The academic revolution*. Garden City, NJ: Doubleday.

Karabel, J. (1984). Status group struggle, organizational interests, and limits of institutional autonomy. *Theory and Society*, *13*, 1–40.

Katz, J. (1978). Epilogue: The admissions process—Society's stake and the individual's interest. In H. S. Sacks (Ed.), *Hurdles: The admission dilemma in American higher education* (pp. 318–347). New York, NY: Atheneum.

Klugman, J. (2012). How resource inequalities among high schools reproduce class advantages in college destinations. *Research in Higher Education*, *53*, 803–830.

Klugman, J. (2013). The advanced placement arms race and the reproduction of educational inequality. *Teachers College Record*, *115*, 1–34.

Kobrin, J. L., Patterson, B. F., Barbuti, S. M., Mattern, K. D., & Shaw, E. J. (2008). *Validity of the SAT for predicting first-year college grade point average*. New York, NY: The College Board.

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

Lewin, T. (2013, August 2). Testing, testing: Most students are taking both the ACT and SAT. *The New York Times*. Retrieved from http://www.nytimes.com/2013/08/04/education/edlife/more-students-are-taking-both-the-act-and-sat.html

McDermott, A. (2008, September 25). Surviving without the SAT. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Surviving-Without-the-SAT/18874

Meredith, M. (2004). Why do universities compete in the rankings game? An empirical analysis of the effects of the U.S. News and World Report college rankings. *Research in Higher Education*, *45*, 443–461.

Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, *1*, 894–904.

Merton, R. K. (1957). *Social theory and social structure*. New York, NY: Free Press.

Park, J. J. (2012). It takes a village (or an ethnic economy): The varying roles of socioeconomic status, religion, and social capital in SAT preparation for Chinese and Korean American students. *American Educational Research Journal*, *49*, 624–650.

Perna, L. W., May, H., Yee, A., Ransom, T., Rodriguez, A., & Fester, R. (2013). Unequal access to rigorous high school curricula: An exploration of the opportunity to benefit from the International Baccalaureate Diploma Programme (IBDP). *Educational Policy*. Advance online publication. doi:10.1177/0895904813492383

Posselt, J. R., Jaquette, O., Bielby, R., & Bastedo, M. N. (2012). Access without equity: Longitudinal analyses of institutional stratification by race and ethnicity, 1972-2004. *American Educational Research Journal*, *49*, 1074–1111.

Robinson, M., & Monks, J. (2005). Making SAT scores optional in selective college admissions: A case study. *Economics of Education Review*, *24*, 393–405.

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, *121*, 297–317.

Royston, P. (1991). Estimating departure from normality. *Statistics in Medicine*, *10*, 1283–1293.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, *135*, 1–22.

Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., Shen, W., & Kiger, T. B. (2012). The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. *Psychological Science*, *23*, 1000–1007.

Shaw, E. J., Kobrin, J. L., Patterson, B. F., & Mattern, K. D. (2012). *The validity of the SAT for predicting cumulative grade point average by college major*. New York, NY: College Board.

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, *32*, 247–261.

Stevens, M. L. (2007). *Creating a class: College admissions and the education of elites*. Cambridge, MA: Harvard University Press.

Supiano, B. (2012, August 17). Emory U. intentionally misreported admissions data, investigation finds. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/blogs/headcount/emory-u-intentionally-misreported-admissions-data-investigation-finds/31215

Thacker, L. (Ed.). (2005). *College unranked: Ending the college admissions frenzy*. Cambridge, MA: Harvard University Press.

Vigdor, J. L., & Clotfelter, C. T. (2003). Retaking the SAT. *The Journal of Human Resources*, *38*, 1–33.

Webster, T. J. (2001). A principle component analysis of the U.S. News and World Report tier rankings of colleges and universities. *Economics of Education Review*, *20*, 235–244.

White, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review*, *21*, 149–170.

Yablon, M. (2001). Test flight: The scam behind SAT bashing. *New Republic*, *30*, 24–25.

Authors

ANDREW S. BELASCO is a PhD candidate in the Institute of Higher Education at the University of Georgia and is also founder and lead consultant at College Transitions Limited Liability Company (LLC). His research focuses on college access and choice.

KELLY O. ROSINGER is a doctoral candidate in the Institute of Higher Education at the University of Georgia. Her research examines efficiency and equity outcomes of higher education at the federal, state, and university level.

JAMES C. HEARN is professor in the Institute of Higher Education, University of Georgia. His research focuses on policy, organization, and finance in post-secondary education.

Manuscript received August 29, 2013
First Revision received February 23, 2014
Second Revision received April 15, 2014
Accepted April 21, 2014

Available online at www.sciencedirect.com
SCIENCE DIRECT
Economics of Education Review 24 (2005) 393–405
Economics of Education Review
www.elsevier.com/locate/econedurev

Making SAT scores optional in selective college admissions: a case study

Michael Robinson[a], James Monks[b,*]

[a]*Department of Economics, Mount Holyoke College, South Hadley, MA 01075, USA*
[b]*Department of Economics, Robins School of Business, University of Richmond, VA 23173, USA*

Received 27 October 2003; received in revised form 19 May 2004; accepted 29 June 2004

Abstract

Despite heightened scrutiny of the use of standardized tests in college admissions, there has been little public empirical analysis of the effects of an optional SAT score submission policy on college admissions. This paper examines the results of the decision by Mount Holyoke College to make SAT scores optional in the admissions process. We find that students who "under-performed" on the SAT relative to their high school GPA were more likely to withhold their scores; the admissions office rated applicants who withheld their scores more highly than they otherwise would have been rated; and, matriculants who withheld their scores had a lower average GPA than those who submitted their standardized test results.

JEL classification: I2; J24

Keywords: School choice; Resource allocation

1. Introduction

The SAT test has come under increasing scrutiny and pressure recently from many fronts. Perhaps the largest challenge to this standardized test, which is widely used in college admissions, has come from Richard Atkinson, President of the University of California System. He has publicly criticized the SAT and the over-reliance of colleges on this test in granting admission to their institutions. He recommends the elimination of the SAT as an admission requirement in the University of California System.[1] Atkinson, and other critics of the SAT, find fault with the test on a number of levels. One criticism is that individuals can be coached to perform well on the test. Another concern is that in the end the SAT does not provide that much additional information concerning the future academic ability of college applicants. This final criticism may be particularly applicable at smaller institutions, where each application is read by an admissions officer, and high school grade point average, class rank, rigor of high school curriculum, and letters of recommendation taken collectively

*Corresponding author. Tel.: +1-804-794-7678; fax: +1-804-794-7680
E-mail addresses: mirobins@mtholyoke.edu (M. Robinson), jmonks@richmond.edu (J. Monks).

[1]There are actually two sets of exams. The SAT I, which is a general test of verbal and quantitative abilities, and the SAT II which are a series of topical exams. Throughout this paper the use of the term SAT refers to the SAT I exam.

0272-7757/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2004.06.006

may provide more than adequate predictive information concerning an applicant's likely success in college.

While the storm brewing in California has garnered a great deal of attention due to the sheer size of the California higher education system and the potential for a rippling effect across the states, it is not the first institutional challenge to the use of the SAT in college admissions. In fact, over 700 colleges and universities provide some form of flexibility in the submission of standardized tests scores in the application process.[2] For example, for over seventeen years Bates has not required standardized tests scores as a criterion of admission (Hiss, 2001). Additionally, Dickinson, Muhlenberg, and Union College, among others, no longer require the SAT or ACT for admission. On the other hand, Lafayette College experimented with optional SAT score submission in admissions, and decided to resume the requirement of submission of a standardized test score for admittance.[3]

There are both potential costs and benefits to an institution of following an optional SAT score submission policy in admissions. One of the potential benefits is that an institution may receive additional applicants, as individuals who otherwise would have chosen not to apply may now do so. This is a benefit to institutions for two reasons. The first reason is that there may be students among the additional applicants with desirable characteristics or qualities that the institution would like to attract. For example, the marginal applicant pool may possess additional minority students and students with outstanding academic characteristics other than SAT scores. A second reason institutions may benefit from an increase in applications is that they appear more selective, as they accept the same number of students from a larger applicant base. A lower acceptance rate is one measure used by magazines, students, and admissions counselors to gauge the academic quality of an institution.

Another potential benefit to an institution in implementing an optional SAT score submission policy is that it may result in a higher reported average SAT. In fact, one observer (Yablon, 2001) called the use of optional SAT score submission a "scam", in which institutions use optional SAT score submission as a means of raising their average reported SAT scores of the entering class. Higher average reported scores make an institution appear more selective and of higher quality. Ehrenberg (2001) questions the motivation of these institutions for making SAT scores optional. He suggests, as does Yablon, that institutions have made this change in policy, at least in part, to bolster their positions in influential rankings such as the *US News and World Report* rankings of colleges. There is prima facie evidence both in support and opposition to this charge. Brownstein (2001) reports that no longer requiring the SAT actually lowered the reported SAT scores of Dickinson and Franklin and Marshall, while raising the average SAT scores of Muhlenberg, despite the fact that Muhlenberg reports the average SAT score over all of their students, including those who blocked their scores during the application process.

On the other hand, there are costs associated with not collecting the SAT scores of all of the applicants. It may be the case that without the SAT scores of a student the admissions office loses an important tool in differentiating the caliber of student they would like to admit from those they would not. In this scenario, admissions officers may respond by assuming that all non-submitters are "lemons" and not admitting any of the students who withhold their SAT scores. Because the admissions process repeats itself every year, and students are able to observe, at least second hand, previous years' outcomes, this behavior is not sustainable. Soon no one would apply without submitting one's scores and the policy would be moot.

From the students' perspective, the important question becomes how they perceive the institution treats applicants who opt not to submit their scores. A student would only withhold her score if she felt it would improve her chances of being admitted, given her SAT score and other academic characteristics. Similarly, a student would only submit her score if she felt it improved her chances of being admitted. Students who feel that they possess attributes that the college would find desirable, but who did not perform well on the SAT, would be more likely to withhold their SAT scores, while students who performed well on the SAT relative to their other academic credentials would be more likely to submit their scores.

These questions are important from a public policy standpoint. The popularity of SAT test-prep courses represents a considerable investment in attempting to increase one's score on the SAT. If institutions are able to differentiate among its applicants without the use of SAT scores, then these resources could be devoted elsewhere. Second, in as much as enrollment to selective institutions is limited, how these scarce positions in the enrolling class are distributed among the applicants is an important question of allocative efficiency. The SAT is intended to be a signal of academic potential. If it is a noisy signal, then individuals may be self-selecting not to apply to certain institutions where they feel, based on their SAT scores, they do not have a reasonable chance of admission. This result may be most acute among racial and socio-economic groups that traditionally do not perform as well on the SAT. Making the SAT optional in admissions may result in a different distribution of the limited enrollment seats.

[2] This figure comes from the Fairtest organization (http://fairtest.org/optinit.htm).

[3] McCarty (2001).

Despite the heightened concern with the use of standardized tests in college admissions and the important institutional and public policy implications of the use of SAT scores in admissions, there has been little public empirical analysis of the effects of an optional SAT score submission policy on admission outcomes and the subsequent academic performance of those students who chose not to submit their scores, but were admitted to the institution. This paper attempts to fill this void by examining the results of the recent decision by Mount Holyoke College, a small, prestigious, New England, women's, liberal arts college, to make SAT score submission optional in the admissions process. The following analyses focus on the effect of the optional SAT score submission policy on: (1) the size and racial composition of the applicant pool; (2) the decision of applicants to either submit or withhold their SAT scores in the admissions process; (3) the treatment of applicants who choose to withhold their SAT scores by the admissions office; (4) the yield (percentage of admitted applicants who matriculate) of test submitters versus non-submitters; and finally (5) the academic performance of the applicants who blocked their scores during the admissions process. The emphases of these analyses are on the individual choice of whether to submit one's scores, and whether the institution can make an informed decision concerning the academic prospects of the candidate for admission without knowing her SAT score.

2. Data

The data for this analysis are taken from the freshmen class entering in the fall of 2001, the first cohort of applicants for whom the new optional SAT score policy was implemented. For the purpose of this study only applicants for whom admissions decisions were made are considered (individuals with incomplete application materials were excluded from the data). In addition, submitters are defined to be all students who submitted and did not block either an SAT or ACT score and non-submitters are defined to be all applicants for whom the admissions office made an admissions decision based on neither a submitted SAT nor ACT score.[4]

There were 2627 applicants for this class versus 2445 applicants for the entering class of 2000 (see Table 1). This represents an increase of approximately 7 percent. In comparison, the median increase in applications across a set of 12 peer institutions was 1 percent.[5] The increase at Mount Holyoke was the result of a 6 percent increase in white applications and a 21 percent increase in minority and international applications, versus 1 percent and 2 percent increases among our peers, respectively. There was a 27.8 percent decline in the number of applications for whom race was unknown, and therefore it is unclear how much of the increase in minority applications is simply due to re-classification.

Table 1
Completed applications by ethnic status

	2000	2001	% increase (%)
White	1134	1202	6.00
Black	126	208	65.10
Hispanic	86	131	52.30
Asian	223	275	23.30
Native American	8	19	137.50
International	532	549	3.20
Race Unknown	336	243	−27.70
Total	2445	2627	7.44

Table 2
Non-submitters vs. submitters sample (all applicants for which decisions were made)

	Percent	*N*
Submitted no score	24.2	637
Submitted any Score	75.8	1990
Submitted only SAT	70.7	1856
Submitted only ACT	2.3	61
Submitted both SAT and ACT	2.8	73
International		
Submitted no score	33.0	181
Submitted any score	67.0	368
Domestic		
Submitted no score	21.9	456
Submitted any score	78.1	1622

Table 2 illustrates that among the 2627 applicants, 24.2 percent chose to block their standardized tests scores from the admissions office. Thirty-three percent of international students did not submit their scores, and 21.9 percent of domestic students did not submit their scores. These results are similar to the percentage of non-submitters reported by Bates College during the

[4]Individuals who submitted ACT scores rather than SAT scores were eliminated from the following analyses in order to avoid problems of accurately converting ACT to comparable SAT scores.

[5]The 12 peer institutions are Amherst, Barnard, Bryn Mawr, Carleton, Oberlin, Pomona, Smith, Swarthmore, Trinity, Wellesley, Wesleyan, and Williams.

Table 3
Admit rate and yield by submit/non-submit of SAT scores

	Applicants	Accepts	Admit rate (%)	Matriculants	Yield (%)
Total	2627	1451	55	507	35
Submitted scores	1990	1169	58	394	34
Did not submit scores	637	282	44	113	40

first five years under their SAT optional policy (Bradley, 1990).

Table 3 provides a comparison of submitters versus non-submitters based on their application status. As mentioned above, 24.2 percent of applicants did not submit a standardized test score, 19.4 percent of admitted applicants did not submit a test score, and 22.3 percent of the ultimate matriculants did not submit a test score. Based on these summary results it does not appear that the institution changed its policy simply to inflate the number of applicants with no intention of admitting them, in order to increase its reported selectivity. Similarly, it does not appear that the admissions staff viewed all non-submitters as "lemons" and thus did not admit them.

Table 4 compares summary measures of submitter versus non-submitter applicants. As expected, the non-submitters have lower SAT scores. Average SAT scores for non-submitters are based on the 48.4 percent of non-submitters for whom we were able to obtain SAT scores from the College Board or from their high school transcripts, after the admissions process was complete. There are no statistically significant differences in the average high school GPA or class rank of the non-submitters for whom we have SAT scores versus those for whom we do not have test scores. This suggests that the sample of non-submitters for whom we have SAT scores is academically comparable to the sample of non-submitters for whom we do not. Among applicants the non-submitters average combined SAT score is 141 points lower than the average combined SAT score for submitters.

The non-submitters also have lower average high school GPA and class rank than submitters. We use the high school GPA as reported by the high school to the college. All data reported on other than a four point scale have been recalibrated to the 4 point scale. (For example, 100 would correspond to 4.00.) Because high schools typically assign a higher weight to advanced placement (AP) courses than regular courses the highest GPA attainable is greater than 4.00. Home schooled applicants and those whose high schools do not report GPA's are omitted from the statistical analyses using GPA. The high school GPA gap between submitters and non-submitters is 0.18. Class rank is reported as the ratio of the students numerical rank in the class divided by the size of the graduating class times 100 (or the percentile rank reported by the high school). The lowest possible rank is therefore 100 and the highest rank is less than 1 for graduating classes of more than 100 students. For the valedictorian the class rank would be $(100)1/N$ where N is the class size. The difference in average class rank between submitters and non-submitters is 4.8.

Table 4
Differences between submitters and non-submitters all applicants

	Submitters	Non-submitters	Significant
Family contribution	$21,742	$19,785	Yes
No-need	33.0%	27.1%	Yes
Math SAT	617	550	Yes
Verbal SAT	633	558	Yes
Total SAT	1250	1109	Yes
High school GPA	3.61	3.43	Yes
Class rank	15.6	20.4	Yes
Fall accept rate	58.7%	44.3%	Yes
Early decision	8.1%	10.2%	No
White	49.4%	34.2%	Yes
Black	6.9%	11.1%	Yes
Hispanic	4.7%	5.8%	Yes
Asian	10.2%	11.5%	No
International	18.5%	28.4%	Yes
Race unknown	9.6%	8.2%	No
Admission rating	4.2	4.8	Yes
No. of observations	1990	637	

SAT scores for non-submitters are based on those non-submitters for whom scores are available (48.4 percent of non-submitters). Significance at the 5 percent level.

It is also interesting to note the differences in non-academic characteristics of submitters versus non-submitters. The non-submitters are less likely to be white. Approximately, thirty-four percent of non-submitters versus 49 percent of submitters are white. Non-submitters are also more likely to be non-US citizens. Twenty-eight percent of non-submitters versus 18.5 percent of submitters are non-US citizens. These results coupled with the significant increase in minority applications relative to white applications outlined in Table 1 suggest that the change in policy may have had a positive impact on racial diversity. Additionally, because income data is limited for non-aid applicants we use the family contribution (FC) estimated by the financial aid office using the Institutional Methodology (IM) as a proxy for income. Family contribution is the amount that the family is expected to contribute towards the total educational cost of the student. For non-aid recipients the FC is equal to tuition, fees, room, and board. The non-submitters have lower average income as reflected in their lower average FC and are less likely to be among the students who do not receive any financial aid (no-need) than submitters.

There are no significant differences in the percentage of submitters versus non-submitters who applied (or were admitted) early decision.

These summary measures suggest a number of factors that may contribute to the decision to submit one's standardized test scores. The following section outlines an empirical model of the applicant's decision to submit one's SAT scores, and the related decision by the institution to admit an applicant given that she did or did not submit her SAT score.

3. Empirical model

Clearly, there are two application pools present here. One pool of applicants submitted their SAT scores, and the other applicant pool did not submit their scores, which application pool a candidate belonged to is obviously endogenously determined. This is an example of a switching regression model, with endogenous switching. The first equation, whether to submit one's SAT scores or not, determines to which application equation one is assigned.

$$S^* = X\beta + \mu \text{ where } S = 1 \text{ if submit,}$$
$$S = 0 \text{ if non-submit,} \tag{1}$$

$$A_1 = Z\gamma_1 + (\text{SAT})\gamma_2 + \varepsilon_1 \text{ if } S = 1, \tag{2}$$

$$A_2 = Z\gamma_3 + \varepsilon_2 \text{ if } S = 0, \tag{3}$$

where Eq. (1) is the dichotomous decision to submit or withhold one's SAT scores , A is a measure of admissions (either admissions rating or the dichotomous variable of admittance or not), Z are non-SAT measures of admissions attributes, such as high school GPA and class rank, and X are factors that determine the decision to submit one's SAT scores.

The first step is to estimate the decision to submit one's SAT scores or not (Eq. (1)). It is expected that an applicant would be more likely to submit her SAT score if she felt that it would improve her chances of being admitted ($A_1 > A_2$). Consequently the vector of factors X in Eq. (1) must include all elements of Z and SAT scores. An applicant is likely to feel that submitting her SAT scores will improve her chances of admission the higher her SAT scores. On the other hand, it is expected that an individual with lower SAT scores would be less likely to submit one's scores. Additionally, conditional on SAT scores an individual with other desirable attributes would be more likely to withhold her SAT scores. Similarly, the greater the weights placed on non-SAT measures in the non-submission equation relative to the weights placed on non-SAT characteristics and SAT scores in the submission equation, $(\gamma_3 - \gamma_1)/\gamma_2$, the less likely one is to submit one's SAT scores.

The error terms in Eqs. (2) and (3) are not mean zero, as individuals self-select which applicant pool to join based on which pool they believe will increase their probability of admittance. To correct for this selectivity, the results of the SAT submission Eq. (1) are used to perform a Heckit correction for self-selection on the sample of applicants that submitted their SAT scores (Eq. (2)). The selectivity corrected coefficients from this regression were then used to predict the admissions rating of the non-submitters. This predicts the admissions rating that would have been given the non-submitters if they had submitted their scores. Next we estimate the selectivity corrected coefficients for non-submitters (Eq. (3)), and apply the coefficients from this regression to the submitters. This now estimates the admissions rating that would have been given the submitters if they had withheld their SAT scores.

As students are not ultimately interested in their admissions ratings per se, but rather in whether they are admitted or not, we also estimate the above system of equations examining the dichotomous admissions result of admittance or not ($A = 1$ if admitted, zero if not). In order to test for the impact of test score submission on admittance, accounting for the self-selectivity of test score submission, a sequential bivariate probit with partial observability ($S = 1$ if they submitted test scores, $S = 0$ if they did not; $A = 1$ if they were admitted, $A = 0$ if they were not) is estimated on the sample of applicants who submitted their SAT scores. The sequential nature of the bivariate probit comes from the fact the applicants first choose whether to submit their SAT scores or not. The partial observability aspect of the bivariate probit stems from the fact that we only observe the admissions decisions based on SAT scores for those individuals who submitted their SAT scores. Similarly, we only observe the admissions decisions without the use of SAT scores for those individuals who withheld their scores. Following the approach called by Meng and Schmidt (1985) partial, partial observability, the likelihood function to be estimated for SAT score submitters is

$$L = \prod_{A_1=1} \frac{F(X\beta, Z_1\gamma_1 + (\text{SAT})\gamma_2, \rho)}{\Phi(X\beta)} \times \prod_{A_1=0} \frac{\Phi(X\beta) - F(X\beta, Z_1\gamma_1 + (\text{SAT})\gamma_2, \rho)}{\Phi(X\beta)}, \tag{4}$$

where Φ is the standard normal distribution, and F is the bivariate normal distribution, and ρ is the correlation of the error term in the submission decision equation with the error term in the acceptance equation, for SAT score submitters.

Following the approach outlined above, the coefficients from this acceptance model are then used to predict the probability of admission of the non-submitters had they been treated like SAT score submitters. The bivariate probits were identified by excluding FC from the acceptance equation, while including it in the submit equation. As a test of robustness the bivariate probits were also performed including FC in both equations and identifying off of functional form. The results are qualitatively the same.

A similar bivariate probit is also performed for non-submitters and the coefficients from this regression are applied to the characteristics of the submitters to estimate the probability that they would have been admitted had they withheld their SAT scores.

Finally, we attempt to ascertain whether the admissions staff is able to accurately differentiate the academic abilities of the applicant pool without the use of SAT scores for all students. If the admissions staff is able to successfully identify the most able students, then conditional on the admissions rating assigned to each individual, whether they withheld their SAT scores or not should not affect either their yield rates or their academic performance once on campus. To test these hypotheses, we first perform a probit of the decision to enroll or not conditional on admissions rating and a dummy variable for withholding one's score. Next we regress the first year grade point averages of the enrolled students on their admissions rating and a dummy variable for withholding one's SAT score. If the admissions staff can accurately assess the academic caliber of students without the aid of SAT scores, then the coefficient on withholding one's SAT scores should not be statistically significant, conditional on one's admissions rating, in determining a student's probability of matriculating or first year grade point average.

4. Empirical results[6]

The summary measures and empirical model outlined above suggest that race, citizenship, one's SAT scores, and family income may be influential in determining whether to submit one's SAT scores for admission to Mount Holyoke College. As discussed above, the lower one's SAT score the less likely she would be to submit her score. Second, the higher the probability the student would be admitted based on other attributes without a SAT score the less likely she would be to submit her scores. This would imply that applicants with higher GPA's and better class rank would, ceteris paribus, be less likely to submit a SAT score. To test these hypotheses a probit model of SAT score submission is estimated.

4.1. The decision to submit one's SAT scores

Table 5 contains the estimated coefficients from the probit regression of SAT score submission. The two main hypotheses are generally supported. Applicants with higher math and verbal SAT scores are more likely to submit their test scores, and the higher the high school GPA of the applicant the less likely she is to submit her scores.[7] There is no significant effect of class rank on the probability of submitting one's SAT scores. Two other interesting results emerge. First, there is no significant effect of race on the probability of submitting scores. Interestingly, the coefficients on the black and Hispanic indicator variables are positive suggesting that, if anything, blacks and Hispanics are more likely to submit scores than whites, ceteris paribus. This appears to contradict the prediction that applicants with other characteristics that are desirable to the institution would be less likely to submit their scores. If there is affirmative action for minority applicants, then one would expect minority applicants to be less likely to submit their scores. On the other hand, a minority applicant may view a given SAT score as being more meritorious relative to her minority peers than the same score as a white applicant, and she may be more inclined to submit her score. Another interesting result is that the more affluent the applicant, as measured by the FC, the less likely one is to submit a score. This result may indicate that applicants from wealthier families either have more a priori confidence that they will be admitted, perhaps because of the quality of the high school they attend. It may also be because a given SAT score may be viewed by an applicant from an affluent area of the country as being less noteworthy, in comparison to her peers, than the same score obtained by an applicant from a less well-to-do area of the country. As a result, the more affluent applicant may be more likely to withhold her test score, while the less affluent applicant submits her score. An alternative explanation is simply that the wealthier

[6] The programs used to generate the results of this paper are available from James Monks. The data are proprietary to Mount Holyoke College and are not available to the public.

[7] Individuals with missing values for any of the regressors were excluded from all of the following analyses. Also excluded from the analyses were individuals who submitted ACT scores rather than SAT scores. Alternative specifications were analyzed that included the average class rank, high school GPA, or family contribution for one's group (submitted, non-submitted) if this value was missing, and a dummy variable for missing value of the regressor. The results are qualitatively the same.

Table 5
Probit model for submitting SAT scores

Variable	Coeff.	*t*-stat.	Significant
Intercept	−3.118	−5.06	**
Family contribution ($000s)	−0.010	−2.37	**
Math SAT	0.005	6.50	**
Verbal SAT	0.004	6.02	**
Black	0.098	0.51	
Hispanic	0.061	0.30	
Asian	−0.084	−0.52	
Native American	0.267	0.45	
International	−0.054	−0.31	
Class rank	−0.001	−0.19	
High school GPA	−0.284	−2.14	**

Number of observations = 1311.
Included among the regressors, but not shown, are dummy variables for region and unknown race. Omitted race category is white.
** (*) indicates significance at the 5 percent (10 percent) level.

applicants may be better informed about the admissions process and perhaps more adept at playing the admissions game and making strategic choices about whether to submit their scores.

The result that the higher the applicant's high school GPA the less like they are to submit an SAT score should be somewhat reassuring as far as the impact of the policy on student quality, because it suggests that it is higher quality applicants as measured by GPA (or presumably other non-quantitative measures of quality) that are less likely to submit scores, conditional on their SAT scores. It appears that individuals who "underperformed" on the SAT based on their high school achievement are those individuals who are more likely to not submit their test scores. To test this hypothesis we regressed combined SAT scores on high school GPA, class rank, and a dummy variable indicating if the individual did not submit her score (results not shown). The mean SAT scores for submitters was 1250 compared to 1109 for non-submitters. The coefficient on the dummy variable was −130, indicating that non-submitters performed 130 points worse on average than their peers with comparable high school GPA and class rank. Clearly, the non-submitters are individuals who on average performed less well on the SAT than would be expected based on their high school performance. As a group they are "poor test takers". This is similar to the result reported by Bradley (1990) for Bates College.

4.2. Treatment of non-submitters in the admissions process

A primary concern to both individuals and the institution is whether the admissions process works differentially for submitters and non-submitters. Table 6 shows the percentage of each admission rating that were non-submitters both among all applicants, those that were accepted, and matriculants. Mount Holyoke College uses an admission rating system that runs from a high of 1 to a low of 8. Applicants with ratings lower than 6 are typically not admitted. The rating is separate from the admission decision so not all of those within an admission rating will be admitted although the higher the rating the higher the probability of acceptance. The ratings are assigned by readers on the basis of their overall impression of the application and then confirmed at a committee meeting by the entire staff. There are no quantitative guidelines for assigning the ratings and non-ability related factors (such as legacy status) that might affect the acceptance decision do not (in principle) affect the rating. There are substantial numbers of non-submitters in all rating groups, though they are more heavily concentrated in the lower rating categories among applicants. Forty five percent of those rated 8 (the lowest rating category) among applicants were non-submitters. This tends to give the impression both that some non-submitters were attempting to manage overall bad profiles by not submitting their scores, and that they were not particularly successful in this attempt. We can get a glimpse of the impact of non-submission on admission rating by comparing the non-submitted and submitted SAT scores in each rating group. Fig. 1 shows this for accepted applicants (and Fig. 1b for all applicants). Two interesting results emerge. First, there remain substantial gaps in the combined SAT score at each admission rating of approximately 100 points, for both applicants and admits. This is consistent with the idea the non-submitters are "poor test takers" and that their scores under predict their other academic credentials. However the pattern of scores across ratings is remarkably similar. The difference in the mean combined score between ratings 1 and 6 is 304 for submitted scores and 254 for non-submitted scores. One interpretation of this result is that the non-submitted scores are highly correlated with other factors used to rate

Table 6
Percent non-submitters by admission rating

Rating	Applicants	Admits	Matriculants
1	11.5	10.2	16.7
2	12.9	12.7	17.7
3	17.8	15.3	17.8
4	23.8	22.6	22.9
5	27.1	28.0	30.3
6	31.4	22.6	18.6
7	34.0	NA	NA
8	45.6	NA	NA

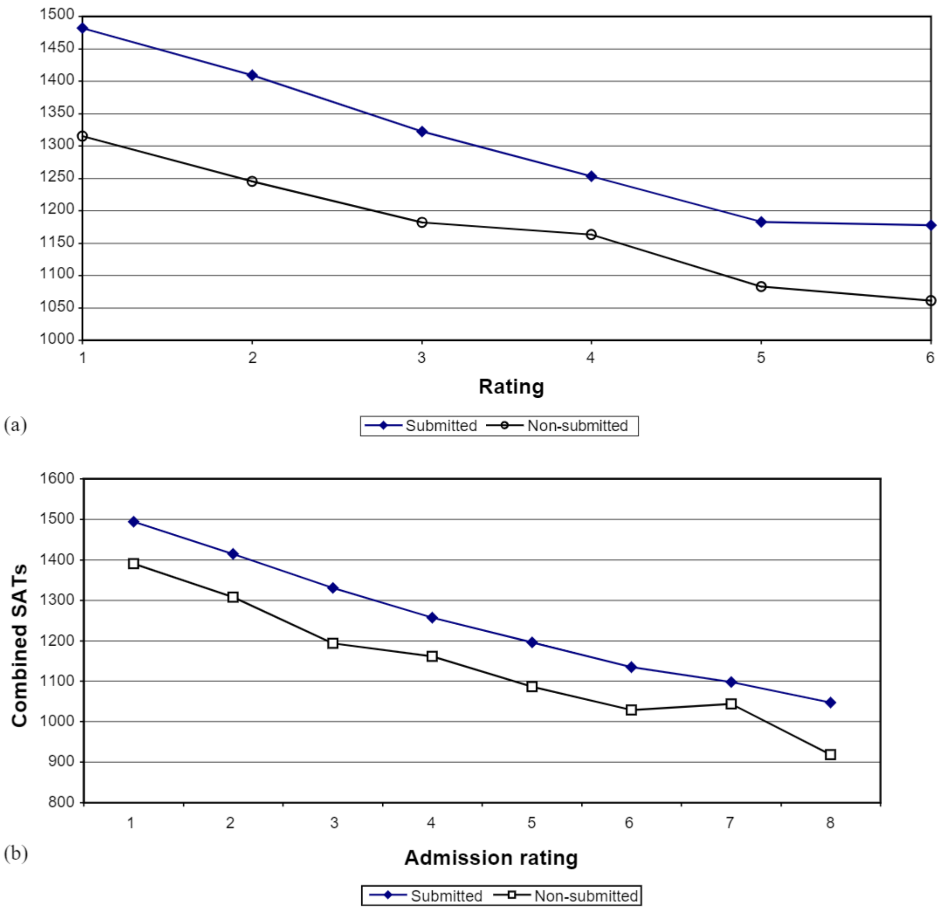


Fig. 1. (a) SAT scores by admission rating. (b) SAT scores by admission rating—all applicants.

applicants, such as high school GPA, class rank, letters of recommendation, and personal essays. This suggests that SAT scores could have been used to separate the non-submitters into their admission ratings, but that it is not necessary to have SAT scores to determine one's appropriate admission rating. The admissions office appears to be able to separate non-submitting applicants into rating groups that reflect the underlying scores, without using the actual SAT scores.

Another test of the ability to rate the students without using the SAT can be obtained by comparing the high school GPA of the submitters and non-submitters by admission rating, among all applicants. Fig. 2 shows this relationship. When examining the overall characteristics of the submitters and non-submitters we observed that on average the non-submitters had lower high school GPA (see Table 4). However, Fig. 2 reveals that there is virtually no difference in high school GPA between the groups within admission ratings. This convergence of GPA's in the presence of a SAT gap is consistent with our evidence the SAT seems to under predict ability as measured by high school GPA for non-submitters and that the final admission rating reflects ability. As a final measure of the effect of the submission or non-submission of scores on admission rating, a Heckit selectivity corrected regression is estimated on the sample of submitters with admission rating as the dependent variable, as outlined above. Included among the regressors are verbal SAT scores, math SAT scores, high school GPA, class rank, and dummy variables for race (black, Hispanic, Asian, Native American, and unknown race), international student, and region.[8] The actual average admission rating of non-submitters was 4.6. Applying the coefficients from this selectivity corrected regression of submitters to the characteristics of the non-submitters yields a predicted average admission rating for non-submitters, had they been treated the

[8]We are prevented from presenting the coefficients of the admissions rating regression and admittance probit due to the proprietary nature of the data. As our primary emphasis is on the treatment of non-submitters in the application process, we focus on their treatment in this process rather than on the influence of the other determinants in the application decision.

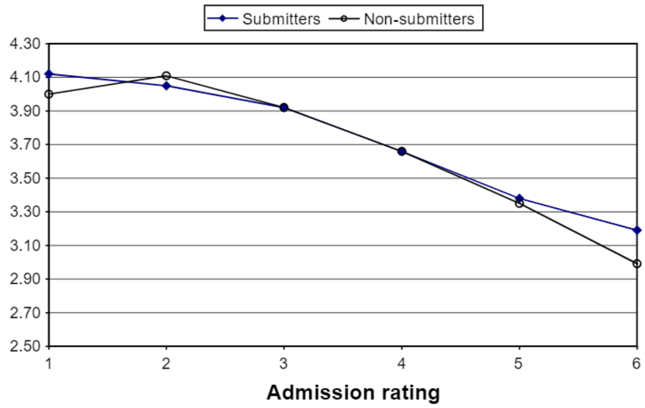


Fig. 2. High school GPA by admission rating.

same as those who submitted their scores, of 5.4 (one is the best admission rating and eight is the worst). The overall difference in the average predicted and actual admission ratings was 0.8 (see Table 7). This suggests one of two possibilities. That the non-submitted SAT scores are poor measures of the applicants' ability and that the admissions office rates these applicants higher based on other information or that there is a slight benefit to not submitting scores. An additional regression of admission rating was performed for test score submitters excluding SAT scores from the regressors. The coefficients from this regression were again used to predict the admissions rating for non-submitters. This predicted value estimates the average admissions rating non-submitters would have received had they been treated the same as their peers with comparable high school GPA and class rank, but who chose to submit their scores. In this case, the predicted average admissions rating for non-submitters was 3.7, indicating that although non-submitters were given better average admissions ratings than their peers with comparable SAT scores, they were given less favorable ratings than their peers with the same high school GPA and class rank.

Similarly, as illustrated in Table 7, using the selectivity corrected coefficients from the admissions rating equation for non-submitters to predict the admissions rating for submitters results in a predicted average rating for submitters of 4.4 versus an actual admission rating of 4.1. As expected, on average submitters would have been worse off had they not submitted their SAT scores.

These results suggest that those individuals who chose not to submit their scores were wise to not submit; however, the admissions office gave these individuals lower average ratings than just their high school GPA and class rank would suggest, in a sense discounting their high school performance for not submitting their SAT scores, but not discounting their admissions rating to the point that is justified by their SAT scores.

After applicants are rated, the decision must be made about whether to accept them or not. We have already observed that the accept rate for non-submitters was substantially lower than for submitters (see Table 4), though we have now seen that non-submitters received slightly higher admission ratings than otherwise comparable submitters, conditional on SAT scores. Because the acceptance decision is a dichotomous choice ($A = 1$ if admitted, $A = 0$ if not admitted), the estimation of the selectivity corrected decision to admit an applicant or not becomes a sequential bivariate probit with partial observability, as outlined above. We estimate this bivariate probit for the sample of SAT score submitters and apply the coefficients from this regression to the characteristics of non-submitters in order to predict their probability of admittance had they been treated the same as the submitters. The actual accept rate for the non-submitters was 52.6 percent, while the predicted accept rate was 38.8 percent (see Table 7).[9] Similar to the results for admission rating, this suggests an advantage towards being accepted for non-submitters, conditional on their SAT scores. Following the approach used to analyze the admissions ratings above, we also estimated the admittance bivariate probit excluding SAT scores from the regressors, and again predicted the accept rate of non-submitters had they been admitted at the same rate as submitters with comparable high school GPA and class rank. The average predicted acceptance rate was 61.7 percent. Fewer non-submitters were actually admitted than would have been the case had they been treated the same as submitters with the same GPA and

[9]The actual admit rate used here varies from the overall admit rate for non-submitters because we only include those individuals for whom we had SAT scores and who reported their high school GPA and class rank.

Table 7
Actual versus predicted admissions outcomes for submitters and non-submitters

	Actual	Predicted	Difference
Submitters			
Admission rating	4.1	4.4	0.3
Percentage admitted	61.9%	61.2%	0.7%
Non-submitters			
Admission rating (incl. SAT)	4.6	5.4	0.8
Admission rating (excl. SAT)	4.6	3.7	−0.9
Percentage admitted (incl. SAT)	52.6%	38.8%	14.8%
Percentage admitted (excl. SAT)	52.6%	61.7%	9.1%

Table 8
Yield by admission rating for submitters and non-submitters Regular decision only

	Submitters		Non-submitters	
Admission rating	*N*	Yield	*N*	Yield
1	53	18.9	6	33.3
2	139	16.6	20	25.0
3	368	17.9	61	19.7
4	267	27.0	76	29.0
5	134	26.1	51	33.3
6	13	69.2	4	25.0
Total	974	22.1	218	27.1

None of the differences (including the overall difference) are statistically significant.

class rank, but more non-submitters were actually admitted than would have been the case had they submitted their SAT scores.

Applying the coefficients of the bivariate probit for non-submitters to the characteristics of submitters resulted in an average predicted accept rate of 61.2 percent versus an actual acceptance rate of 61.9 percent. Submitters were slightly better off having submitted their test scores than would have been the case had they not submitted.[10]

4.3. The matriculation rates of non-submitters

Once accepted, the applicant chooses whether or not to enroll. The yield for non-submitters (40.1 percent) is higher than the yield for submitters (33.1 percent). There may be several factors at work here. One possibility is that because many colleges require SAT scores the non-submitters are either disadvantaged in their applications to other institutions and not accepted, or they choose not to apply to other institutions. In either case, we would expect them to yield at higher rates than submitters. Another explanation is lower student quality. We have already seen that the non-submitters are weaker applicants than the submitters as measured by high school GPA, class rank, and admission rating. This would also lead to higher yields. Table 8 shows yields by admission ratings for submitters and non-submitters. Among regular decision fall admits (thus excluding early decision admitted students) the non-submitters have a yield of 27.8 percent compared to 22.1 percent for submitters (though this difference is not significant). However, within admission ratings the differences were usually quite small and never statistically significant. The biggest differences were among admits rated 1 and 2, where non-submitters yielded at 33.3 percent and 25.0 percent, and submitters yielded at 18.9 percent and 16.6 percent, respectively, although these differences are not statistically significant due to the small cell sizes within these admission ratings. In admission rating 6 there is also a large (but insignificant) difference in yield but with cell sizes that are so small that there is little information here.

To further explore these hypotheses a probit model of the probability of enrolling was estimated. Here the focus is on the probability that non-submitters will matriculate, conditional on having been accepted, and on admission rating. The dependent variable is one if the individual enrolls and zero if they do not. Conditional on the admission rating of the accepted candidate, the coefficient on the dummy variable for not submitting one's standardized tests scores is positive, but not significantly different from zero. There do not appear to be any significant yield differences between submitters and non-submitters conditional on having been accepted and admissions rating.

These results suggest a number of potential conclusions. Since it was not the case that the non-submitters yielded at higher rates after controlling for admission rating, it does not seem likely that they were disadvantaged in their applications to other institutions. This might mean that other institutions while requiring the SAT do not use it too heavily in the admissions process. It also may suggest that the admissions office was able to place the applicants into admissions rating categories that were appropriate given their overall quality and choice set of competing schools even without direct knowledge of their SAT scores.

[10] All of the above results on the actual versus predicted acceptance rate of submitters and non-submitters are qualitatively the same when using just univariate probits that do not account for self-selection into the different applicant pools.

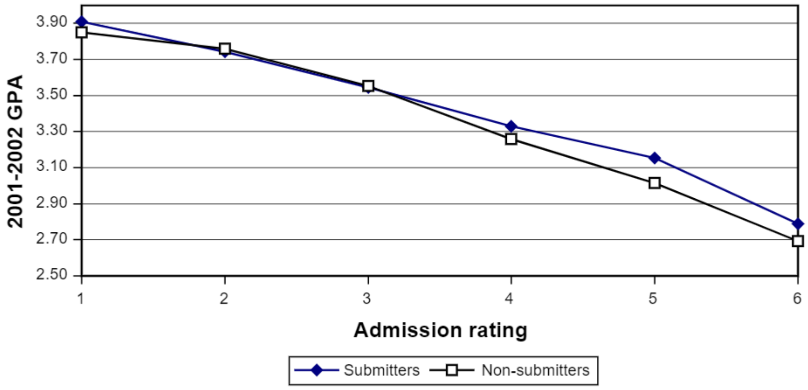


Fig. 3. First year GPA by admision rating for submitters and non-submitters.

4.4. The academic performance of non-submitter matriculants

Overall the non-submitters had a slightly lower first year GPA (3.24) than the submitters (3.35). Of course this could be expected given the lower overall admission ratings of the non-submitters and their lower high school GPAs. Fig. 3 presents relationship between first-year GPA and admission rating for the submitters and non-submitters. For those in admission ratings 2 and 3 the non-submitters had higher first year GPAs than the submitters and overall the admission rating seems to map well against first-year GPA.

In order to further test this relationship Table 9 presents the results of a regression of first-year GPA against FC, variables to measure difficulty of schedule (percent of courses in the humanities, percent of courses taken in math/science, percent of courses taken at the 200 or 300 level), dummy variables for admissions rating, race, international student, region, early decision, and a dummy variable equal to one if the student did not submit her SAT score. The coefficients on the admissions rating dummy variables are positive and statistically significant, indicating that individuals who are rated more highly in the admissions process do indeed perform better academically. Of primary concern here is the coefficient on the non-submission dummy variable. Conditional on one's admissions rating, and demographic characteristics non-submitters performed .08 points worse on average than those who submitted their SAT scores. This is statistically significant at the 10 percent level. Because it appears in Fig. 3 that most of the difference between the GPA of submitters and non-submitters occurs at the higher admissions ratings, we estimated the model with separate non-submission dummy variables, one for students who had admission ratings 1-3 and one for those rated 4–6. In this model the coefficient for those rated 4–6 is larger (−0.14) and significant at the 5 percent level.

Table 9
Grade point average regression analysis Dependent variable = -first year GPA

	Coef.	T-stat.	Coef.	T-stat.
Intercept	2.647	19.63**	2.658	19.66**
Non-submitter	−0.083	−1.94**		
Non-submitter (AR1-AR3)			−0.050	−0.93
Non-submitter (AR4-AR6)			−0.142	−1.99**
Percent humanities	0.280	2.51**	0.281	2.52**
Percent science	−0.236	−2.15**	−0.243	−2.21**
Percent 200–300 level	0.010	0.15	0.009	0.13
Admission rating 1	1.102	8.45**	1.086	8.28**
Admission rating 2	0.936	10.07**	0.919	9.75**
Admission rating 3	0.746	10.81**	0.730	10.31**
Admission rating 4	0.539	8.16**	0.521	7.64**
Admission rating 5	0.349	5.02**	0.357	5.11**
Black	−0.028	−0.34	−0.031	−0.37
Asian	0.024	0.38	0.016	0.25
Hispanic	−0.131	−1.53	−0.136	−1.59
Race unknown	−0.056	−1.36	−0.059	−1.35
International	0.129	−1.35	0.128	−1.30
Family contribution ($000)	0.002	−0.92	0.002	−0.97
Early decision	0.033	0.85	0.034	0.85

Note: ** (*) indicates significance at the 5 (10) percent level. Also included among the regressors but not shown are regional dummy variables.

4.5. The impact of the optional SAT policy on average reported SAT

Ignored until this point has been the impact that this policy has had on the reported measures of academic quality of the institution. Table 10 shows changes in the most commonly reported measures of academic quality for the class entering in the fall of 2000 (the year before the change in policy) to the class entering in the fall of 2001 (the year after the change in policy). While year-to-year class differences should not be solely attributed to a

Table 10
Comparison of the class entering 2000 with the class entering 2001

	Applicants			Admits			Matriculants		
	2000	2001	2001 (incl. non-submits)	2000	2001	2001 (incl. non-submits)	2000	2001	2001 (incl. non-submits)
Math SAT	611	617	608	629	636	628	613	616	607
Verbal SAT	630	634	624	656	658	648	646	643	630
Total SAT	1243	1252	1232	1285	1295	1277	1259	1259	1237
GPA	3.55	3.57		3.69	3.74		3.59	3.64	
Rank	16.4	16.7		11.5	11.9		14.2	14.7	
White	46.4	45.8		52.6	47.9		55.01	52.9	
Black	5.6	7.9		4.6	8.8		3.4	5.1	
Hispanic	3.5	5		4.3	6.5		2.8	4.9	
Asian	9.1	10.5		11.8	13.8		10.1	10.1	
International	21.8	20.9		11.8	12.1		13.8	16.6	
Race unknown	13.7	9.3		14.5	10		14.2	9.5	

single policy change, they do provide a benchmark for assessing the policy. Table 10 outlines the differences in the characteristics of the applicants, admitted students and matriculants from 2000 to 2001. For all three groups, high school GPA is slightly higher, while class rank is slightly worse in 2001 versus the 2000. Submitted SAT scores increase for applicants and accepts, while remaining unchanged for matriculants. Total SAT scores, including both submitted and non-submitted declined for all three groups. Additionally, there appears to be a slight increase in the application and matriculation of minority students. These results are more consistent with the possibility that the change in policy encouraged more applications from individuals with low SAT scores, than that the policy encouraged those with low scores who would have applied anyway to not submit their scores. The end result is that average reported SAT scores did not increase at Mount Holyoke in the first year after the implementation of the optional SAT policy.

5. Conclusion

The use of SAT I scores in college admissions has come under heightened scrutiny of late. This analysis attempted to examine whether selective college admissions could be successfully performed without the requirement of standardized test scores from all applicants. The primary conclusion from this analysis is that selective college admissions can indeed be carried out under an optional SAT score submission policy at an institution. These results stem from a case study at a single institution and thus represent a partial equilibrium in that the ability to successfully identify and admit students without the use of SAT scores may be dramatically different and more difficult if all institutions adopted this policy and continued to follow it over a long period of time. In the case of Mount Holyoke College, however, it appears that some of the potential benefits of the optional SAT policy may have been achieved. There is some evidence of an increase in applications relative to a set of peer institutions, and there may have been an increase in minority applications, as well. On the other hand, the change in policy did not result in an increase in the average reported SAT. At the same time, it appears that these benefits did come at some costs. The students who withheld their SAT scores and ultimately were admitted and enrolled had a lower average GPA than their peers with comparable admissions ratings but who submitted their SAT scores. So the benefits discussed above were tempered by the loss in information that may have been garnered from the SAT scores of all of the applicants.

It remains to be seen if these preliminary results remain consistent over time. It may be the case that as applicants become better informed about the optional SAT policy at Mount Holyoke, the percentage and profile of applicants choosing to not submit their scores will change rendering the admissions office task more difficult. The experience at Bates College, however, suggests that this will not be the case. Additionally, the long run success of non-submitters in terms of cumulative GPA, graduation rates, and satisfaction of these students with their educational experience is yet to be determined and warrants future examination. The long run impact of this policy on the composition of the student body and the overall academic quality of the institution is also fertile ground for future analysis.

While this analysis suggests that there are both benefits and costs to an institution of pursuing an optional SAT admissions policy these results may not be universally applicable. Mount Holyoke is a small liberal arts college with less than 3000 applications. The time and resources it is able to devote to each individual application may not be available at larger institutions with many more applications. Additionally, if more institutions follow this admissions approach the impact on the admissions environment would appear to be minimal. The yield on non-submitters, conditional on admissions rating, is not significantly different from the yield on submitters suggesting that Mount Holyoke College does not appear to be treating these individuals substantially differently than most other comparable institutions. Should more institutions pursue this policy it is not apparent that the admissions and enrollment decisions would be dramatically different.

References

Bradley, D. R. (1990). Committee on admissions and financial aid. The optional SAT policy at Bates: A final report. February. Lewiston, Maine: Bates College.

Brownstein, A. (2001). Colleges debate whether dropping the SAT makes them more competitive. Chronicle of Higher Education. October 26.

Ehrenberg, R. G. (2001). Reaching for the brass ring: How the US News and World Report rankings shape the competitive environment in US higher education. Cornell Higher Education Working Papers, no. 17.

Hiss, W. (2001). Optional SAT's and Bates: Seventeen years and counting. Chronicle of Higher Education. October 26.

McCarty, B. (2001). Lafayette's comfort level is higher with SAT's. Chronicle of Higher Education. October 26.

Meng, C., & Schmidt, P. (1985). On the cost of partial observability in the bivariate probit model. *International Economic Review*, *26*(1), 71–85.

Yablon, M. (2001). Test flight: The scam behind SAT bashing. *New Republic*, *30*, 24–25.

Economics Letters 174 (2019) 13–17

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

SAT optional policies: Do they influence graduate quality, selectivity or diversity?

Matt Saboe [a,*], Sabrina Terrizzi [b]

[a] West Chester University, 700 S. High St, West Chester, PA 19383, United States
[b] Moravian College, 1200 Main St. Bethlehem, PA 18018, United States

HIGHLIGHTS

- SAT optional policies have no effect on racial and socioeconomic diversity.
- SAT optional policies do not influence the gender ratio of institutions.
- SAT optional policies have no effect on the quality of the student population.
- SAT optional policies briefly increase applications, but the effect is not sustained over time.

ARTICLE INFO

Article history:
Received 1 October 2017
Received in revised form 27 September 2018
Accepted 16 October 2018
Available online 22 October 2018

JEL classification:
I23

Keywords:
SAT optional
Selectivity
Higher education
Enrollment quality

ABSTRACT

Despite many conversations regarding the applicability and relevance of the SAT as a valid admissions tool, there is limited evidence regarding the effects of test-optional policies on various aspects of an institution's effectiveness and the collegiate experiences within each institution. Using data from the Integrated Postsecondary Education Data System (IPEDS) coupled with a difference-in-difference analysis, we find that test-optional policies have very limited effects. We find SAT optional policies to have no significant effect on diversity or enrolled student quality. The only statistically significant effect we find is a brief increase in the number of applicants in response to the new policy.

1. Introduction

Despite many conversations regarding the applicability and relevance of the SAT as a valid admissions tool (Robinson and Monks, 2005), there is limited evidence regarding the effects of test-optional policies on various aspects of an institution's effectiveness and the collegiate experiences within each institution (Belasco et al., 2014; Robinson and Monks, 2005). The few studies addressing this topic highlight the lack of additional information provided by the SAT in the college admissions process (Belasco et al., 2014; Robinson and Monks, 2005); however, they still identify the SAT as a predictor of collegiate success when coupled with high school GPA (Barbuti et al., 2008; Marcus, 1989; Sackett et al., 2012; Shaw et al., 2012). Amid growing concerns of the lack of predictive power of the SAT for minority students (Hoffman and Lowitzki, 2005; Nettles et al., 2003) and its inherent bias (Grodsky et al., 2008), some institutions of higher education are choosing to eliminate the testing requirement as a means of consideration in the application process. However, there exists a fear of the potential negative effects to quality, enrollment, and retention if SAT scores are not used in the admissions process (Belasco et al., 2014; Robinson and Monks, 2005).

In this analysis, we add to the growing literature on the effects of test-optional policies on various institutional outcome measures. Specifically, we assess the test-optional policy's effectiveness on retention and graduation rates across all four-year, not-for-profit, baccalaureate-granting institutions during the 2009 to 2014 time period. We find that adopting a test-optional policy leads to a short-lived increase in applications, but has no statistically significant effect on other measures of quality and selectivity. Similarly, we find test-optional policies have no effect on diversity.

2. Literature review

The transition to test-optional policies began in 1964 at Bowdoin College in Maine (Lewis, 2015). Bates College followed twenty years later with the adoption of their test-optional policy in 1984 (Hiss, 2001). Since the 1980s and 1990s, this movement has spread

* Corresponding author.
E-mail addresses: msaboe@wcupa.edu (M. Saboe), terrizzis@moravian.edu (S. Terrizzi).

https://doi.org/10.1016/j.econlet.2018.10.017
0165-1765/

Table 1
Summary statistics outcome variables.

Variable	Test required		Test optional	
	Mean	Std. Dev.	Mean	Std. Dev.
Retention rate	73.73	13.38	78.12	17.23
Graduation rate	51.57	19.56	61.05	22.59
Admission yield	36.63	17.27	40.69	26.36
Applications	5753.00	8226.71	2851.75	3246.72
SAT math 75th percentile	589.12	71.22	627.66	66.83
SAT verb 75th percentile	581.95	69.74	624.77	70.17
% Minority	0.39	0.23	0.38	0.21
% Receiving pell grant	39.78	18.52	33.50	23.71
% Male applicants	0.42	0.14	0.43	0.21
Student expenditure FTE	7.79	0.75	8.09	1.03
Instructional expenditure FTE	8.98	0.52	9.24	0.65
Endowment FTE	9.07	1.75	9.76	2.33
Student-to-faculty ratio	14.90	4.72	11.28	5.88
Ln tuition	9.57	0.78	10.10	0.65
Ln enrollment	7.77	1.30	6.89	1.20
% Receiving any financial Aid	89.40	13.30	84.82	19.00

Note: Summary statistics are based on 9658 test-required colleges and 520 test-optional colleges.

across all types of institutions of higher education. In 2015, an additional 47 colleges instituted test-optional admissions policies, increasing the total number of institutions[1] with test-optional policies to approximately 850 (Simon, 2015).

The reasons for implementing test-optional policies vary. Critics argue that test-optional policies simply provide a means of increasing the applicant pool in order to raise the perceived selectivity of the school, while also increasing the reported SAT scores of each incoming class (Belasco et al., 2014; Robinson and Monks, 2005). Contrastingly, proponents claim that test-optional policies expand diversity by increasing access for students of high quality with poor test scores, who tend to be disproportionately of lower socioeconomic status or a racial minority (Alon and Tienda, 2007; Hoffman and Lowitzki, 2005).

The current literature offers few insights into the effects of test-optional policies on the quality of an institution's enrollees, their success within the institution, or the diversity within the institution. In the most robust analysis to date, Belasco et al. (2014) consider the effects of SAT optional policies within a subset of institutions: selective liberal arts colleges. Belasco et al. (2014) find that test-optional policies increase the perceived selectivity of these institutions by increasing applications and reported test scores. These authors do not find evidence of increased diversity as measured by the percent of Pell grant recipients or the proportion of minority students enrolled in these institutions (Belasco et al., 2014).

3. Data and methodology

We use data from the Integrated Postsecondary Education Data System (IPEDS), which are collected by the National Center for Education Statistics. We construct a panel from 2009 to 2014[2] for all four-year, public and private, not-for-profit, baccalaureate-granting institutions of higher education. The sample consists of 1649 (93 percent) colleges requiring SAT/ACT scores and 127 (7 percent) colleges with a test-optional policy over the time period of our study.

Our empirical strategy is a panel difference-in-difference approach to isolate the effect of test-optional policies while removing any unobserved time-varying effects. Our econometric approach is given by:

$$Y_{it} = \alpha_0 + \beta_1 I_i + \beta_2 T_t + \beta_3 Optional_{it} + \gamma X_{it} + \varepsilon_{it}. \quad (1)$$

The coefficient, β_3, of our key variable, $Optional_{it}$, measures the effect of the institution's test-optional policy on our outcome variables. Our outcome variables include measures of student quality and institutional selectivity: six-year graduation rate, full-time student retention rate, admission yield, number of applicants, and 75th percentile SAT scores for both math and verbal sections. We also include measures of diversity: percent minority (non-white) enrolled, percent Pell grant recipients, and percent of applicants who are male. Table 1 shows the summary statistics for these outcome variables. We include institution and time fixed effects, I_i and T_t, to control for any bias as a result of the varying implementation of these policies (Belasco et al., 2014; Bertrand et al., 2004).[3]

Finally, we include a vector, X_{it}, to control for institutional characteristics including: student service expenses,[4] instructional expenses,[5] and endowment per full-time equivalent enrollment. We also include the student-faculty ratio, the natural logarithm of full-time enrollment, and the natural logarithm of tuition and fees. Table 1 shows the summary statistics of these covariates. We use robust standard errors clustered at the college level (Stock and Watson, 2008).

Because we are testing multiple hypotheses simultaneously, i.e. the significance of the SAT optional policy on the multiple outcome variables described above, our p-values from Eq. (1) may be underestimated (Romano and Wolf, 2005). Thus, in addition to the standard errors calculated within each individual regression, we calculate the Romano–Wolf adjusted p-values using the RWOLF command in Stata (Clarke, 2016). This procedure calculates Romano and Wolf's stepdown adjusted p-values, which are corrected for multiple hypotheses testing. Full details of the procedure are described in Romano and Wolf (2016).

4. Results and discussion

Table 2A displays the effects of test-optional policies on various measures of student quality and institutional selectivity. We find a

[1] All types of institutions, including: two-year, four-year, for-profit, and not-for-profit.

[2] We begin our data analysis in the year 2009 because of changes made to variables in 2008. Beginning in 2009 provides us with the use of consistently defined and collected institutional characteristics.

[3] Time fixed effects are jointly significant after we conduct a Wald test.

[4] A functional expense category that includes expenses for admissions, registrar activities, and activities whose primary purpose is to contribute to students' emotional and physical well-being and to their intellectual, cultural, and social development outside the context of the formal instructional program.

[5] A functional expense category that includes expenses of the colleges, schools, departments, and other instructional divisions of the institution and expenses for departmental research and public service that are not separately budgeted.

Table 2A
Test-optional effect on student quality and selectivity.

	Retention rate	Graduation rate	Admission yield	Applications	SAT math 75th percentile	SAT verb 75th percentile
Test-optional policy	2.558	1.284	−2.374**	32.819	−9.816*	−8.534
	(1.62)	(1.48)	(1.16)	(96.17)	(5.94)	(6.39)
Student expenditure	−0.211	−0.548	−0.680	107.273	0.080	−0.167
	(0.78)	(0.76)	(0.97)	(67.20)	(2.64)	(2.73)
Instructional expenditure	0.057	0.846	1.668	573.256***	−0.028	8.452*
	(1.41)	(1.22)	(1.39)	(145.25)	(4.50)	(4.48)
Endowment	0.196	−0.040	0.260	12.834	−0.791	−1.257
	(0.20)	(0.23)	(0.21)	(40.53)	(0.74)	(2.06)
Student-to-faculty ratio	0.115	−0.025	0.163*	48.008***	−0.345	−0.454*
	(0.08)	(0.07)	(0.08)	(11.40)	(0.23)	(0.24)
Tuition	−0.918	−0.612	0.589	7.624	4.112***	0.428
	(0.80)	(0.50)	(0.56)	(136.63)	(1.37)	(1.12)
Enrollment	1.377	2.186	3.115**	1,979.153***	5.923	4.216
	(1.62)	(1.44)	(1.49)	(221.30)	(4.69)	(4.62)
% Financial Aid	0.013	0.054*	0.001	−9.209**	−0.068	0.006
	(0.04)	(0.03)	(0.03)	(3.66)	(0.08)	(0.09)
Constant	68.179***	33.098**	−4.312	−16,500.63***	521.705***	490.045***
	(21.80)	(16.33)	(19.37)	(2849.25)	(58.43)	(60.26)

Robust standard errors are clustered at the college-level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
All columns report difference-in-difference coefficient estimates for 8868 observations. All specifications include time and college fixed effects.

Table 2B
RWOLF P values for effect of test-optional policy on student quality and selectivity.

	Retention rate	Graduation rate	Admission yield	Applications	SAT math 75th percentile	SAT verb 75th percentile
Original P value	0.115	0.385	0.040**	0.733	0.099*	0.182
RWOLF P value	0.436	0.832	0.297	0.941	0.436	0.644

Robust standard errors are clustered at the college-level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
All columns report difference-in-difference p values and Romano Wolf corrected p values for statistical significance of SAT Optional Adoption treatment over 8868 observations. All specifications include time and college fixed effects.

Table 2C
Time passage relative to year of adoption of test-optional policy: Student quality and selectivity.

	Retention rate	Graduation rate	Admission yield	Applications	SAT math 75th percentile	SAT verb 75th percentile
2 Year lead	2.217	6.295**	−1.675	471.470***	5.906	4.046
	(2.54)	(2.99)	(1.60)	(172.09)	(4.38)	(5.93)
1 Year lead	0.112	0.182	−0.298	249.101	3.002	0.205
	(2.66)	(1.91)	(1.43)	(153.22)	(4.39)	(5.38)
Year of adoption	1.741	−1.317	−1.489	398.048***	−9.721***	−7.082*
	(1.23)	(1.10)	(1.32)	(111.30)	(3.75)	(3.79)
1 Year lag	1.786	3.049**	−2.516*	324.538***	−3.041	−6.842
	(1.17)	(1.49)	(1.37)	(96.46)	(2.82)	(4.40)
2 Year lag	2.999***	0.412	−2.080**	120.021	−2.807	−3.895
	(1.06)	(1.25)	(1.00)	(82.39)	(2.86)	(3.12)

Robust standard errors are clustered at the college-level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
All columns report difference-in-difference coefficient estimates for 8868 observations. All specifications include time and college fixed effects. Regressions include controls variables for student expenditure, instructional expenditure, endowment, student-to-faculty ratio, tuition, enrollment, and percent receiving financial aid.

statistically significant effect of test-optional policies on admission yield and the reported 75th percentile math SAT score. These initial results indicate that transitioning to a test-optional policy will make a college appear less selective through reduced admissions yield and lower average SAT math scores.

However, when we use the Romano–Wolf adjusted p-values, we no longer find a significant effect of the SAT optional policy on admission yield and the reported 75th percentile math SAT scores. All other p-values remain insignificant at the 10% level, as shown in Table 2B, which compares the p-values on our estimated coefficient of interest using the standard panel regression models to the Romano–Wolf adjusted p-values. Thus, the initial effects are overstated, and we cannot conclude that SAT optional policies have a significant effect on any of our measured outcome variables.

Table 2C presents the estimates of a similar model where the treatment effect has been replaced with leads and lags to identify the treatment effect before and after adoption of the policy. From this model we can verify that the treatment and control groups maintain common trends preceding the adoption of the policy, while also examining the treatment effect in the years following the policy change. F tests confirm that the coefficients on the lead variables are jointly insignificant and do not violate the common trends assumption; however, the two-year lead for graduation rate and applications are individually significant.[6]

The estimates from this model reveal several interesting findings. While retention rates did not appear to be affected by the policy, there seems to be a positive, albeit lagged, effect on retention rates. Similar to Belasco et al. (2014), we find that colleges appear to attract more applicants in the short run. However, in our analysis adopters experience a delayed decline in the admissions yield after adopting a test-optional policy. This estimated effect is in contrast to the increased selectivity predicted by Belasco et al. (2014). Moreover, colleges that adopt test-optional policies may experience lower SAT math scores rather than the higher scores predicted by Robinson and Monks (2005), but the effect only occurs

6 Although we cannot estimate the Romano–Wolf p-values for multiple independent variables, i.e. the lead and lag variables, we estimate these values separately. The 1-year lag effect on applications remains statistically significant after adjusting the p-values.

Table 3A
Test-optional effect on institutional diversity.

	% Minority	% Receiving pell grant	% Male applicants
Test-optional policy	−0.016	−0.155	−0.012*
	(0.02)	(0.90)	(0.01)
Student expenditure	0.003	−0.482	0.006**
	(0.00)	(0.45)	(0.00)
Instructional expenditure	−0.006	0.268	−0.010
	(0.01)	(1.20)	(0.01)
Endowment	0.001	0.136	−0.003**
	(0.00)	(0.17)	(0.00)
Student-to-faculty ratio	−0.001	0.089	0.000
	(0.00)	(0.07)	(0.01)
Tuition	0.004	−1.659***	0.005**
	(0.00)	(0.40)	(0.00)
Enrollment	0.072***	0.297	0.018*
	(0.02)	(1.30)	(0.01)
% Financial aid	−0.000	0.435***	0.000
	(0.00)	(0.03)	(0.00)
Constant	−0.124	6.269	0.298***
	(0.20)	(15.70)	(0.11)

Robust standard errors are clustered at the college-level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
All columns report difference-in-difference coefficient estimates for 8868 observations. All specifications include time and college fixed effects.

Table 3B
RWOLF P values for effect of test-optional policy on institutional diversity.

	% Minority	% Receiving pell grant	% Male applicants
Original P value	0.367	0.864	0.081*
RWOLF P value	0.832	0.941	0.436

Robust standard errors are clustered at the college-level in parentheses *** p<0.01, ** p<0.05, * p<0.1.
All columns report difference-in-difference p values and Romano Wolf corrected p values for statistical significance of SAT Optional Adoption treatment over 8868 observations. All specifications include time and college fixed effects.

Table 3C
Time passage relative to year of adoption of test-optional policy: Institutional diversity.

	% Minority	% Receiving pell grant	% Male applicants
2 Year lead	−0.009	−1.263	−0.029
	(0.02)	(2.23)	(0.02)
1 Year lead	0.033	−1.366	0.004
	(0.04)	(1.90)	(0.01)
Year of adoption	−0.016	1.632	−0.003
	(0.02)	(1.33)	(0.01)
1 Year lag	−0.009	−0.823	−0.007
	(0.01)	(1.16)	(0.01)
2 Year lag	0.011	−0.011	0.012
	(0.01)	(1.16)	(0.01)

Robust standard errors are clustered at the college-level in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
All columns report difference-in-difference coefficient estimates for 8868 observations. All specifications include time and college fixed effects. Regressions include controls variables for student expenditure, instructional expenditure, endowment, student-to-faculty ratio, tuition, enrollment, and percent receiving financial aid.

in the year of the policy change, does not persist, and is sensitive to p-value adjustments for testing multiple hypotheses concurrently. In summary, test-optional policies may result in colleges finding better-fitting students, who are more likely to be retained out of the greater number of applicants, who previously did not apply. Unfortunately, selectivity of the institution may be compromised as measured by lower admission yields and SAT math scores.

Table 3A details estimates of Eq. (1) where the dependent variables measure diversity. We find that this policy has no significant effect on the percentage of enrolled students who are a racial minority nor the percentage of students receiving Pell grants. The test-optional policy is marginally significant with a p-value of 0.081 when considering the percentage of male applicants, but this significance disappears when using the Romano–Wolf adjusted p-values (adjusted p-value is 0.436). The original p-values and Romano Wolf adjusted p-values for all estimated outcomes related to diversity are displayed in Table 3B.

In Table 3C, we provide estimates from a similar lead and lag model for the college diversity measures. F tests again confirm common trends in the policy leads eliminating concern of pre-treatment trends biasing the results. These estimates reaffirm that a test-optional policy is not an effective way to encourage diversity.

5. Conclusions

Despite the many claims that test-optional policies will increase diversity, we find no statistically significant evidence to support this hypothesis. Further, we find no evidence that test-optional policies affect quality and selectivity. Using the Romano–Wolf adjusted p-values, SAT optional policies have no effect on any of our measured outcomes. We find only limited evidence that retention rates improve and the number of applications increase, while selectivity suffers through lower admission yields and decreased reported SAT math scores.

We emphasize that these latter results should be interpreted with caution and may warrant further investigation. Within these models, only future applications, the one-year lag in particular, was statistically significant. Thus, it may be the case that the SAT optional policy is a pure marketing strategy resulting in a short-term increase in applications without fundamentally changing the quality or diversity of the institution.

Acknowledgment

We would like to recognize initial work on previous iterations of this paper by Mark Wunderly during his honors thesis at Moravian College.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

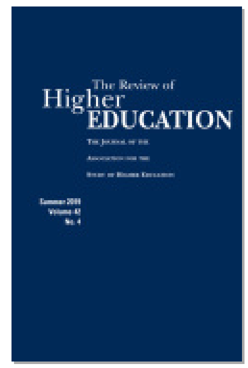
Conflicts of interest

None.

References

Alon, S., Tienda, M., 2007. Diversity, opportunity, and the shifting meritocracy in higher education. Am. Sociol. Rev. 72 (4), 487–511.
Barbuti, S.M., Kobrin, J.L., Patterson, B.F., Shaw, E.J., Mattern, K.D., 2008. Validity of the SAT for Predicting First-Year College Grade Point Average.
Belasco, A.S., Rosinger, K.O., Hearn, J.C., 2014. The test-optional movement at America's selective liberal arts colleges a boon for equity or something else? Educ. Eval. Policy Anal. 0162373714537350.
Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Q. J. Econ. 119, 249–275.
Clarke, D., 2016. "RWOLF: Stata module to calculate Romano-Wolf stepdown p-values for multiple hypothesis testing, " Statistical Software Components S458276, Boston College Department of Economics, revised 06 Jun 2018.
Grodsky, E., Warren, J.R., Felts, E., 2008. Testing and social stratification in American education. Annu. Rev. Sociol. 34, 385–404.
Hiss, W.C., 2001. Optional SATs at Bates: 17 years and not counting. Chronicle Higher Educ. 48, B10–11.
Hoffman, J.L., Lowitzki, K.E., 2005. Predicting college success with high school grades and test scores: Limitations for minority students. Rev. Higher Educ. 28 (4), 455–474.
Lewis, D., 2015. The Ins, Outs of Applying to Test-Optional Colleges. A growing number of colleges are choosing not to require SAT or ACT results. Retrieved February 07, 2016, from http://www.usnews.com/education/best-colleges/articles/2015/09/11/the-ins-outs-of-applying-to-test-optional-colleges.
Marcus, R.D., 1989. Freshmen retention rates at US private colleges: Results from aggregate data. J. Econ. Soc. Meas. 15 (1), 37–55.
Nettles, M.T., Millett, C.M., Ready, D.D., Ludwig, J., Forman, J., Millet, C.M., 2003. Attacking the African American: White Achievement Gap on College Admissions Tests. Brookings papers on education policy, (6), 215-252.
Robinson, M., Monks, J., 2005. Making SAT scores optional in selective college admissions: a case study. Econ. Educ. Rev. 24 (4), 393–405.
Romano, J.P., Wolf, M., 2005. Exact and approximate stepdown methods for multiple hypothesis testing. J. Amer. Statist. Assoc. 100 (469), 94–108.
Romano, J.P., Wolf, M., 2016. Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. Statist. Probab. Lett. 113, 38–40.
Sackett, P.R., Kuncel, N.R., Beatty, A.S., Rigdon, J.L., Shen, W., Kiger, T.B., 2012. The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. Psychol. Sci. 0956797612438732.
Shaw, E.J., Kobrin, J.L., Patterson, B.F., Mattern, K.D., 2012. The validity of the SAT for predicting cumulative grade point average by college major.
Simon, C.C., 2015. The Test-Optional Surge. Retrieved February 07, 2016, from http://www.nytimes.com/2015/11/01/education/edlife/the-test-optional-surge.html?_r=0.
Stock, J., Watson, M., 2008. Heteroskedasticity-robust standard errors for fixed effects panel data regression. Econometrica 76 (1), 155–174.

Test-Flexible Admissions Policies and Student Enrollment Demographics: Examining a Public Research University



Paul G. Rubin, Manuel S. González Canché

The Review of Higher Education, Volume 42, Number 4, Summer 2019, pp. 1337-1371 (Article)

Published by Johns Hopkins University Press
DOI: *https://doi.org/10.1353/rhe.2019.0068*

The Review of Higher Education
Summer 2019, Volume 42, No. 4, pp. 1337–1371
 (ISSN 0162–5748)

Test-Flexible Admissions Policies and Student Enrollment Demographics: Examining a Public Research University

Paul G. Rubin and Manuel S. González Canché

Abstract: A growing number of postsecondary institutions in the United States have removed standardized testing as a requirement for admission. Researchers, however, have suggested that these "test-optional" policies may not benefit underrepresented populations as intended, but instead serve as an additional

---

*Correspondence to: Paul Rubin, University of Utah, 1721 Campus Center Drive, Room 2220, Salt Lake City, UT 84112. Email: *paul.rubin@utah.edu*

Paul G. Rubin is a postdoctoral research associate at the University of Utah's Department of Educational Leadership and Policy. Paul's research agenda is broadly focused on the intersections between higher education policy, governance of the postsecondary sector, and the use of academic research in the policy process. His recent work considers how state and institutional contexts influence policy decisions around college access and completion.

Manuel S. González Canché. I hold a Ph.D. in higher education with cognates in Biostatistics, Sociology, and Economics. I'm currently an associate professor at the University of Pennsylvania's Higher Education Division. I'm particularly interested in and committed to understanding structural factors that influence minoritized, low-income, and "at-risk" students' likelihood of educational and occupational success. Relying primarily on geographical network analysis and quasi-experimental methods design as analytic tools, I seek to find more nuanced understandings of the effect of students' choices with respect to their location and their college sector of initial attendance on access, persistence, and success in higher education.

revenue source for the institution. In this study, we utilize a synthetic control method to extend this research by considering whether a more nuanced "test-flexible" policy, which allows qualifying students to decide whether to submit test scores instead of an institution-wide policy, influences student enrollment demographics at a public university.

INTRODUCTION

Test-flexible admissions policies are a subset of the broader "test-optional policy" agenda, which seeks to deemphasize the use of standardized entrance exams (e.g., SAT and ACT) for admission to postsecondary colleges and universities in the United States. However, whereas test-optional policies remove standardized test scores universally for all applicants, test-flexible policies are unique in that only applicants who meet specific requirements may apply without submitting test scores (Belasco, Rosinger, & Hearn, 2014; Syverson, 2007). Examples of criteria requirements to qualify under test-flexible policies include a minimum high school grade point average and/or ranking among the graduating high school class, as well as involvement in extracurricular activities and service experiences. Test-flexible policies also tend to allow qualifying applicants to ultimately decide whether to submit their standardized test scores, which provides the student the opportunity to decide whether their score is representative of their abilities or not. Considering that both test-optional and test-flexible policies share the purpose of waiving the often controversial standardized testing requirement, a motivating goal for institutions adopting these policies often centers on increasing opportunities for college access among students who tend to underperform on these high stakes exams.

In particular, despite standardized admissions testing's original intention to provide a "'common currency' that allows admissions officers… to place students on the same footing" (Garvey, 1981, p. 1) when applying to college, there have been concerns of bias against student populations traditionally underrepresented in higher education. Specifically, previous research suggests that students from traditionally underrepresented racial and ethnic populations in higher education (e.g., Black, Hispanic, Native American, and Pacific Islander students), first generation college students, and students from lower socioeconomic levels tend to attain lower scores on standardized tests than their majority and more affluent peers (Atkinson, 2001; Blau, Moller, & Jones, 2004; Camara & Schmidt, 1999; Crouse & Trusheim, 1988; Fleming, 2002; Freedle, 2003; Geiser & Studley, 2002; Hoffman & Lowitzki, 2005; Jencks, 1998; Zwick, 2002; Zwick & Green, 2007). Some have attributed this discrepancy to an unequal access to resources, such as costly preparatory courses and private tutors, and more fundamental exam-level factors, including the vocabulary used and phrasing of questions (Balf, 2014; Kapor & Klein, 2007; Rosner, 2012; Soares, 2012; Zwick, 2004). From this view, test performance does not measure scholastic aptitude, which motivated the initial implementation of the SAT (Garvey, 1981), but rather serves as a magnifier of systematized biases within society. These factors have led many researchers and campus administrators to question the objectivity and reliability of standardized tests as a predictor of college preparedness and success and ultimately challenge its utility in the college admissions process.

Accordingly, an increasing number of colleges and universities have elected to deemphasize testing as a metric for admission through aforementioned test-optional policies. Although test-optional policies differ by institution, colleges and universities often discuss a common goal guiding their decision to deemphasize standardized tests: to improve racial and socioeconomic diversity among enrolling students. Institutions argue that becoming test-optional will allow for a more holistic review of applicants, which enables admissions officials to consider less quantifiable aspects of a student's background (e.g., race, gender, geography, educational opportunity availability, extracurricular activities, etc.) when rendering admissions decisions. From this view, a holistic admissions process may translate into a more diverse and well-rounded student population.

While empirical studies often find students from underrepresented groups are more likely to apply and receive admission to test-optional institutions (Bates College, 2004; Belasco et al., 2014; Hiss & Frank, 2014; Syverson, 2007), some suggest these findings alone do not support the notion that these types of policies are motivated by the goal of improving higher education diversity. For example, studies have discussed concerns about prestige and national ranking contributing to an institution's willingness to become test-optional. In particular, following adoption of these policies, college and universities often observe increases in the number of applications received from students who are deemed under-qualified for acceptance, regardless of background. In turn, institutions can lower their acceptance rate and appear more competitive nationally by rejecting these applicants, who would have otherwise not applied (Ehrenberg, 2002; Epstein, 2009). These studies also suggest the potential for selection bias among students who provide strong standardized test scores to test-optional institutions because it can skew the reported average to be higher than had all enrolled students submitted their test scores. Ultimately, this perspective suggests increased student diversity serves as an unintended consequence of test-optional policies, at best, rather than a direct result, and warrants further examination on its direct impact on institutional student enrollment demographics.

The extant literature investigating the impact of test-optional policies has also had a limited focus. The majority of studies have focused on private institutions, which tend to have greater financial resources available to provide more flexibility in enrolling a class of students to align with institutional goals (Breneman, Doti, & Lapovsky, 2001; Hearn, 2001). On the other hand, while public colleges and universities might have less financial resources available, they are uniquely committed to serving resident students and positioned to influence the enrollment of traditionally underrepresented populations in higher education (González Canché, 2014). Finally, among test-optional policy research, there has been minimal consideration of test-flexible policies, which limit the influence of institutional goals on admissions decisions by allowing qualified students to decide whether to submit test scores. The present study aims to start filling these gaps in the literature by examining the impact of a test-flexible policy on the demographics of enrolled students at one of the first public universities to adopt such a policy in the United States, George Mason University.

Study Purpose

The purpose of this study is to analyze the effect of the implementation of a test-flexible policy at a selective public university on student demographic composition. Considering the controversial role of standardized admissions tests, this purpose constitutes a marked and relevant departure from the current literature on college access and stratification. Specifically, this study is timely and relevant as it enables assessment of the effects of becoming test-flexible on enrollment behaviors, which may potentially lead to increased access among students traditionally underrepresented in higher education. In line with this purpose, this study addresses the following three research questions:

(1) Does the adoption of a test-flexible policy impact the enrollment of low-income students (as measured by Pell Grant recipient enrollment)?
(2) Does the adoption of a test-flexible policy impact the enrollment of students who identify as a member of a traditionally underrepresented minority group in higher education?
(3) Does the adoption of a test-flexible policy coincide with changes to the average amount of institutional grant funding per first-time, full-time student?

These questions were evaluated using a dataset built from official information gathered from the Integrated Postsecondary Education Data System (IPEDS) and the United States Department of Education. The models were estimated using a synthetic control method (SCM), which compares the true results observed at the test-flexible institution with the outcomes of a counterfactual version of the same institution that did not adopt the policy as explained in the methods section.

Literature Review

The interest in adopting test-optional policies by colleges and universities can be traced to two events: a speech by Richard Atkinson, former president of the University of California system (UC), at the 2001 American Council on Education annual meeting and a 2004 presentation at the National Association for College Admissions Counseling (NACAC) national conference by Bates College's Dean of Admissions William Hiss. Atkinson's (2001) address recommended UC move away from requiring standardized test scores for admission and suggested that the "overemphasis on the SAT is distorting education priorities and practices." From his perspective, the SAT did not adequately gauge academic achievement but instead focused on aptitude, which disproportionately undermined minority students and those unable to afford preparation courses. Although UC never dropped standardized testing requirements for admission (Epstein, 2009), Atkinson's speech is considered a contributing factor that led College Board to the 2005 SAT redesign that aimed to focus more on a student's academic preparation than reasoning skills.[1] However, while Atkinson and UC were considering moving away from the SAT as an admissions metric in 2001, Bates College had already been test optional for seventeen years.

In October 1984, the faculty at Bates voted to make standardized testing optional for all applicants, which was overseen by Dean of Admissions William Hiss (Epstein, 2009). Although it was not the first selective institution to adopt a test-optional policy, Bates became an often-cited example due to their multiple evaluation studies post-adoption. Hiss delivered one of the more notable presentations for Bates at the 2004 NACAC annual conference, where he highlighted twenty years of data showing "no differences in academic performance or graduation rates between submitters and non-submitters" (Bates College, 2004; Syverson, 2007) under the institution's test-optional policy. Hiss and Valerie Frank (2014) extended these findings by studying 33 colleges and universities[2] and found similar results regarding academic performance and completion rates between submitters and non-submitters.

Within this broader study, Hiss and Franks (2014) discussed the demographics and backgrounds of students that tend to be non-submitters. They found these students are more likely to be a first-generation college enrollee, a member of a racial minority, a recipient of a Pell Grant, and have a learning disability than those opting to submit standardized test scores. Nevertheless, Hiss and Franks mentioned the broader appeal of test-optional policies not-

[1]The 2005 changes to the SAT made the writing section mandatory and removed analogies and quantitative-comparison problems (Epstein, 2009).

[2]The Hiss and Franks (2014) sample included twenty private colleges and universities, six public universities, five minority-serving institutions, and two arts institutions.

ing white students "use optional testing policies at rates within low single digits of the averages" (p. 3). They also noted a bimodal curve regarding family financial capacity of non-submitters with large pools of students that do not qualify for financial aid taking advantage of test-optional policies, effectively balancing the large proportion of Pell recipients. The wide range of students taking advantage of test-optional policies has contributed to the growing interest nationally, though several researchers have questioned the institutional rationale behind eliminating consideration of standardized tests.

Several studies suggested colleges and universities are motivated to adopt test-optional policies in order to increase admissions selectivity and institutional prestige. For example, Ehrenberg (2002) noted that a criterion in the student selectivity category of the *U.S. News & World Report* (USNWR) ranking is average standardized test scores. Following adoption of a test-optional policy, he argued, "Only students who score well will report [their scores]… [so] average test scores for admitted applicants that are reported to *USNWR* [will rise]" (p. 156). Considering the prominence of USNWR's ranking system, Ehrenberg suggested institutions would consider enacting these policies strictly with the end goal of increasing their rankings in this influential publication. Similarly, studies noted institutions received an increased number of applications following the adoption of test-optional policies from students "who might otherwise not apply" (Belasco et al., 2014, p. 209; Epstein, 2009; Robinson & Monks, 2005). These researchers argued that institutions would be able to deny more of these underprepared applicants and raise admissions selectivity for UNSWR and other ranking systems. Ultimately, these studies suggested institutions might be adopting test-optional policies for self-serving reasons rather than the altruistic messages that are espoused.

Moreover, an understudied outcome of test-optional policies is how their enactment impacts institutional diversity among students enrolled. While we have an understanding of the demographics of those individuals who opt to take advantage of these policies, less is known about the extent that these students ultimately matriculate at a given institution and if these changes are the result of the enactment of the policy or due to other circumstances. A recent study that began to fill this gap in the literature is by Belasco et al. (2014). These researchers examined the impact of test-optional policies on the proportion of low-income and minority students enrolling at selective liberal arts colleges. Using a quasi-experimental, difference-in-differences approach, they found that the enactment of fully test-optional policies, where standardized test scores are not a required component for admissions consideration,[3] has had a non-significant impact in expanding educational

[3]Belasco et al. (2014) noted that at fully test-optional institutions, many students still submit standardized test scores, but, based on the policy, is not considered when determining whether a student is admitted.

opportunity for these underrepresented populations. Further, they concluded that test-optional policies "may perpetuate stratification within the postsecondary sector… by assigning greater importance to credentials that are more accessible to advantaged populations" (p. 218), thereby emphasizing differences among applicants based on their socioeconomic level and/or availability of opportunities.

Although the Belasco et al. (2014) study lays the groundwork for understanding the influence of test-optional admissions on student enrollment demographics, its findings may not hold true for all institutions with such policies. For instance, Syverson (2007) noted great variation in selectivity among institutions included on FairTest's list of test-optional institutions. He highlighted a "substantial portion of them are 'non-competitive' or 'minimally competitive' institutions… [with about one-third] identifying themselves as 'moderately selective'" (p. 62). Syverson also noted differences between policies, with some institutions requesting exam scores "only for placement… [or] giving applicants the option not to have their test results considered in the admission process" (p. 62–63), instead of uniformly removing the standardized testing requirement for admissions consideration.

The current study aims to contribute to the literature by considering Syverson's overview of the test-optional policy landscape and investigating an institution's test-flexible policy, which allows qualified students to decide whether or not to submit standardized test scores. We hypothesize that by providing students a choice, rather than uniformly removing the testing requirement, may result in a more pronounced impact on enrollment demographics than the non-effect reported by Belasco et al. (2014). Further, by considering a public university, we also fill a gap in the literature regarding institutional type. Our study is informed by the previous literature regarding variable selection, including institutional characteristics that may help predict the utility of the test-flexible policy and the development of the conceptual framing that underlies the project.

Conceptual Framework

Theories of college access and choice and academic capitalism guide this study and highlight the perspectives of the two key stakeholders involved in test-optional policies: students and higher education institutions. Throughout the extensive literature on college access and choice, standardized tests are often mentioned as a factor considered by students when they are creating their college choice set. Manski and Wise (1983) suggested that students consider their potential for admission at a given institution prior to applying and select institutions with published standardized test scores similar to their own. Zemsky and Oedel (1983) also mentioned that students consider a more limited geographic range and quality of institution as their standardized test scores and income-level fall, which can negatively affect their college choice options. Furthermore, in discussing their three-phase model, Hossler and Gallagher (1987) argued, "If there are no colleges which meet [a student's] expectations for a desirable college and which fall within a student's localized search, some students may select non-college options" (p. 214). Therefore, test-optional policies should remove a significant obstacle for students when they are developing their college choice set and, in theory, expand potential avenues for college access.

In potential conflict with the student access and choice literature are institutional realities underscored by academic capitalism. This theory argues that higher education institutions are moving towards the pursuit of market and market-like behaviors in order to secure external streams of revenue (Slaughter & Cantwell, 2012; Slaughter & Leslie, 1997; Slaughter & Rhoades, 2004). Slaughter and Rhoades (2004) noted sector-wide shifts from a "public good knowledge/learning regime" to an "academic capitalist knowledge/learning regime," resulting from the blurring roles of markets, state actors, and the higher education sector. Although these researchers emphasized that colleges and universities have not completely replaced their traditional focus on the public good with academic capitalism, there is the potential for institutions to struggle balancing these two somewhat disparate ends.

Among public postsecondary institutions, a potential tipping point is the decreasing financial support for higher education via state funds. Although Hearn (2006) noted that tuition and fees and government appropriations accounted for only "about half of all revenues in public four-year institutions" (p. 28), the decline in state assistance has required institutions to consider alternative means to recover lost revenue. To this end, one potential solution could be the adoption of a test-optional policy. Considering previous studies noted the increase in applications (Bates College, 2004; Belasco et al., 2014; Epstein, 2009; Robinson & Monks, 2005), institutions could focus on admitting more high-socioeconomic students, who are able to pay sticker price without the support of institutions funds, to fulfill the financial gap. Therefore, while an institution may experience an increase in racial diversity following the adoption of a test-optional policy, socioeconomic diversity may remain unimproved, effectively supporting academic capitalism and a movement away from focusing on the public good of all citizens regardless of socioeconomic class.

The convergence of the college access and choice and academic capitalism theories provides an important lens to consider the findings from this study's analysis. Test-optional policies most directly affect student behavior and influence their college decision processes, but the policy itself takes place within a given institution's strategic plan. However, due to the continued loss of monetary resources once provided by state support, institutions must increasingly find new and unique ways to generate revenues to offset these losses. This is especially critical among public universities including the focus of our study, George Mason University.

Study Setting: George Mason University

George Mason University (GMU) introduced a test-flexible policy starting with the class beginning Fall 2007 (GMU, 2006), which made it one of the first selective public institutions to adopt any form of test-optional policy following the 2005 changes to the SAT. GMU's policy allows students who earned a "competitive," cumulative high school grade point average in line with their averages (3.3–3.9 on a 4.0 scale) with "strong performance in a challenging academic curriculum by having taken a robust selection of college preparatory, honors, advanced placement and International Baccalaureate courses" and "strong leadership, motivation and intellectual curiosity should be demonstrated in your extracurricular, work, or service experiences" to be a competitive applicant without submitting standardized test scores (GMU, 2017).[4] The decision to submit or withhold test scores is ultimately at the student's discretion, even if they fulfill these requirements. In fact, GMU (2017) emphasizes, "some students may not wish to submit standardized test scores as a component of the application process because they may believe the SAT or ACT test scores do not adequately reflect their level of academic achievement and/or predict their potential." Because of their early adoption and the structure of their policy, GMU provides a notable case to consider the possible impact of test-flexible policies on student enrollment demographics.

Notably, GMU's stated rationale to enact a test-flexible policy focused on a desire to enroll the most academically prepared student body, rather than goals of increasing diversity. Accordingly, the analytic approach implemented in this study is not limited to merely test whether a policy change rendered a given expected outcome, but rather it tests for viable, yet unplanned shifts in diversity and socioeconomic composition of study body. In this view, an understanding of the change to GMU's student enrollment demographics following the adoption of their test-flexible policy may serve to add a new perspective to the growing literature on this subject.

Data and Variables

We collected a panel data set from IPEDS and the United States Department of Education that ranged from 2004–2015[5] for GMU and the

[4]Certain applicants, including home-schooled students, prospective computer science or engineering majors, and prospective NCAA athletes are unable to qualify for the test-flexible policy (GMU, 2017).

[5]Due to data availability for select outcome variables, the time period for Models 1 and 5 only account for the 2004–2014 time periods.

institutions contributing to the donor pool for the control unit. The data set included institution-level data on the primary outcomes of interest including the percentage of enrolled students receiving a Pell Grant, the percentage of first-time, degree-seeking, undergraduate students identifying as an underrepresented minority (Black, Hispanic, Native American, or Pacific Islander), and average amount of institutional grant aid awarded per first-time, full-time enrolled student. In order to further evaluate if the policy had a greater effect on specific racial groups, we also disaggregated the student racial data to percentage of Black and percentage of Hispanic students within the population enrolled.

The selection of predictor and control variables was guided by the inclusion of indicators utilized in the available literature on the topic (Belasco et al., 2014; Robinson & Monks, 2005). These include markers of institutional selectivity, including average reported SAT score (25th percentile of the combined critical reading and math sections),[6] percentage of applicants admitted, and percentage admitted students yielded. As discussed below, the use of the synthetic control method also prompted the incorporation of predictor variables that are relevant to the prediction of the dependent variables of interest and facilitated the identification of a counterfactual unit through SCM's weighting process. These predictor variables included the size of each institution's first-time, degree-seeking, undergraduate cohort, total price for in-state students living on campus, and state appropriations per full-time enrolled student. Our dataset does not include individual student data, which limits our ability to assess differences between standardized test score submitters and non-submitters, but our use of SCM can provide additional institutional-wide inferences on the impact of the policy adoption.

METHOD

We employ a synthetic control method (SCM) for this study. An extension of the quasi-experimental difference-in-differences (DD) framework, SCM aims to compare treated and control units in a pre-treatment period and a post-treatment period in order to understand the impact of an intervention on the treated unit. As a primary challenge of DD is finding an appropriate control unit unaffected by the treatment, SCM creates an artificially constructed ("synthetic") counterfactual of the treated unit unexposed by the intervention through a weighting process of various members in a larger donor pool. It is expected that a combination of units can provide a better comparison for the unit exposed to the intervention than any single unit alone. Motivated by the work of Abadie and Gardeazabal (2003) and Aba-

[6]For institutions where SAT scores were unavailable, we converted the 25th percentile of the composite ACT score.

die, Diamond, and Hainmueller (2010, 2011), SCM uses the control group's observed outcome to approximate the outcome for the treated unit in the absence of treatment. In this view, the method creates a "clone" of the treated unit from the characteristics of entities in the control group. Ultimately, SCM's inclusion of a combination of several members in a control unit rather than DD's use of a single comparison unit often offers a better comparison between treated and control units.

A synthetic control is a counterfactual unit resulting from a weighted average of information donated by available control units on variables and indicators of interest. SCM makes explicit: (1) the relative contribution of each control unit to the counterfactual of interest and (2) the similarities (or lack thereof) between the unit affected by the event or intervention of interest and the synthetic control, in terms of pre-intervention outcomes and other predictors of post-intervention outcomes. SCM assigns weights between 0 and 1 to each member of the donor pool based on various predictor variables relevant to the outcome, such that the weights sum to 1 and match those characteristics in the treated unit as close as mathematically possible (Abadie et al., 2010; Abadie & Gardeazabal, 2003; Klasik, 2013). The individual donor unit with the highest value has empirically contributed the most toward the generation of the synthetic unit given a combination of the similitude in predictors and the particular outcomes of interest (see Table 1, which identifies in bold the institutions that most resembled GMU for different outcomes). The SCM weighting process ensures the treated and control units are as similar as possible in the pre-treatment period and can, therefore, predict the outcome of the treated unit in the absence of the policy or intervention.

This rationale is mathematically expressed as follows:

$$\alpha_{it} = Y^{I}_{it} - Y^{N}_{it}, \qquad (1)$$

Where is the result of interest and represents the expected difference in outcomes observed between the performance of the treated and control units over time. represents the outcome that would be observed for unit *i* at time *t* if unit *i* (GMU) was not exposed to the policy change, while represents the observed outcome reported by GMU. Given that is not observable, the purpose of SCM is to build this potential or counter-factual outcome using information provided by other institutions that were not exposed to this policy change.

Essentially SCM aims to rewrite equation (1) as follows:

$$\alpha_{1t} = Y_{1t} - \Sigma^{J+1} w^{*}_{j} Y_{jt}, \qquad (2)$$

Where $Y_{1t} = Y^{I}_{it}$ and $\Sigma^{J+1} w^{*}_{j} Y_{jt} \approx Y^{N}_{it}$, the weight w^{*}_{j} applied to each one of the donor units *j* is selected through a data driven approach to find the solution that renders a synthetic control unit that best approximates the unit exposed to the intervention with respect to the outcome predictors (for more details about the optimization methods available see Abadie et al., 2011).

Following the creation of the synthetic control unit, SCM compares the outcomes for the treated and control units in the pre- and post-treatment periods akin to DD. The synthetic control unit is considered a suitable comparison to the treated unit if the difference between the two is statistically insignificant during the pre-treatment period, which signals the treated and control units are structurally similar. Provided this is the case, any statistically significant variation in the post-treatment period can be attributed to the intervention.

Placebo Tests

The main assumption in using SCM is that variation in the treated group's outcome variable is directly the result of the policy intervention.[7] From this perspective, it is expected the synthetic unit that was not exposed to the policy change should show no variation in the outcome variable. For example, if a policy was implemented to increase labor force participation for women with children, any positive change in participation should only be seen after the implementation of the policy if it rendered the expected outcomes. In the case of the control group, the labor force participation of the counterfactual population (e.g., women without children) should not change beyond expected trends after the policy implementation since they were not the group affected by the policy. On the other hand, if the treated and control units change with similar magnitudes and directions such changes cannot be attributed to the policy implementation. The SCM method builds upon this rationale and states that if one takes a control unit that is similar to the treated unit and applies the analytic procedure, outcomes should not change with a similar magnitude and direction following policy implementation since only the treated unit actually experienced the intervention.

SCM placebo tests are conducted by identifying a donor unit (UNITj), which never experienced the policy intervention and contributed the most to the creation of the synthetic control unit of the treated unit (UNITi). To this end, UNITj resembled the treated UNITi the most when predicting variation of the outcome variable given the set of predictor and control variables included in the analyses. In practice, this means that for each outcome variable, the contribution of donor units may differ in the creation of a synthetic control unit. For example, for this study, the University of Rhode Island (URI) was the institution that resembled GMU the most when predicting the variation of percent of Pell Grant recipients enrolled (model 1), percent of Black students enrolled (model 3), and Average Institutional

[7]This is also the primary assumption for the difference-in-differences approach.

Grant aid per FT student (model 5) as a function of the predictors used (see Table 1 for all models fitted). This result indicates that for the estimation of a placebo test, analysts should treat URI as if it had experienced the policy intervention in 2007 and then create a synthetic copy of it using all remaining donor institutions, except GMU (given that GMU implemented the policy) to assess the extent to which these results differ from the findings obtained from the GMU models. Given that URI did not implement the test-flexible policy, outcome variation should present no change before and after the policy implementation when compared to its synthetic version (Abadie et al., 2010). Similarly, for model 2, which uses percent of underrepresented minority students enrolled as the dependent variable, Western Michigan University contributed the most to the creation of the synthetic GMU and was used in the placebo test for this outcome variable. In the case of model 4, which uses percent of Hispanic students enrolled as the dependent variable, Florida State University was used for the placebo test.

Random Permutation Tests

Testing for statistical significance between the treated and control units' outcomes presents a challenge in SCM due to the relatively small number of data points typically available following policy implementation (Abadie et al., 2010). Given that this number is usually below 30, the use of tests that assume normality, such as t-test (Casella & Berger, 2002), is not recommended. One reliable and increasingly popular approach to address this methodological challenge consists of constructing a sampling distribution based on the observed data points, rather than assuming a given distribution *a-priori*. In this paper, we relied on the use of random permutation tests (Phipson & Smyth, 2010), which is sometimes referred to as a randomization test, to test for significant differences between as shown in equation (1) and its SCM form as shown in equation (2).

Random permutation tests rely on resampling the observed data to test the null hypothesis that the distribution of the observations across groups of interests is the same by shuffling the observed data for the treated and control units randomly *x* number of times (Good, 2000).[8] Every time this random shuffle, or permutation, occurs, the means differences between treated and

[8]Random permutation tests are similar to the more widely utilized bootstrapping test in that both rely on resampling the observed data (see González Canché (2019) for an example using random permutation tests in quadratic assignment procedures for network analysis of qualitative data). However, whereas bootstrapping is primarily used to construct confidence intervals and calculating standard errors, permutation tests aim to gauge whether the observed difference is more extreme than the differences across a randomized configurations of the observed data. Considering the current study's interest in testing whether policy implementation resulted in different outcomes between GMU and a synthetic counterfactual, random permutation tests are more appropriate to test findings (Good, 2000).

control units is recorded (González Canché, 2019). This procedure enables the recreation of a sampling distribution where mean differences of a given sample size is recorded *x* number of times. Once these thousands of iterations are completed, we can measure the number of times that these random mean differences were greater in magnitude than the actual mean difference between the un-shuffled treated and control unit. For example, if analysts conduct 10,000 random permutation tests and the random mean differences were higher than the actual mean difference only five times, then they can conclude that out of 10,000 random samples only 0.0005 percent of the time this result was due to random chance (wherein 0.0005 it the results of the ratio 5/10,000).

For the current study, the permutation tests between the outcomes reflected by GMU and its synthetic counterparts were conducted using 10,000 random iterations. Accordingly, the results will indicate that the actual observed differences between GMU and its synthetic control happened by random chance x% out of 10,000 random samples. Note that random permutation tests can be conducted post- and pre-policy implementation, which is the approach we implemented in this study as shown in Figure 1. In the pre-policy implementation, we tested whether differences between the actual and weighted outcomes were statistically significantly different from zero or have a more negative outcome than expected. If the outcomes were zero, this would have indicated that baseline equivalence was satisfied using SCM. If the outcomes were more negative for GMU, this would indicate that GMU would have started at a disadvantage and that the policy may potentially have reversed that issue. For the post-policy implementation, if significant differences are observed using the random permutation test, analysts can conclude that the policy was associated with a significant change in counterfactual outcomes. Appendix Table A shows the results from these tests and the mean differences as indicated in Figure 1. The probability values in Table A correspond to the ratio of the number of times the random differences were as or more extreme than the differences shown in the Table.

Figure 1 also provides clarity regarding the steps followed in the random permutation approach conducted in the study and shows that the tests were conducted with pre- and post-outcomes data. This strategy implies that, due to data availability, for the pre-implementation data we only have access to six observations in total, as shown in the green histogram. This restricted number of data points was randomly reconfigured 10,000 times and the differences of each reconfiguration rendered a distribution with a mean quite close to zero. The observed difference in the pre-treatment period (-.016) indicates that GMU had a lower representation than its synthetic control and the random permutation test with 10,000 iterations indicated that this difference is far enough from zero. Indeed, only in 171 iterations the randomly generated means differences were as or more extreme than the observed

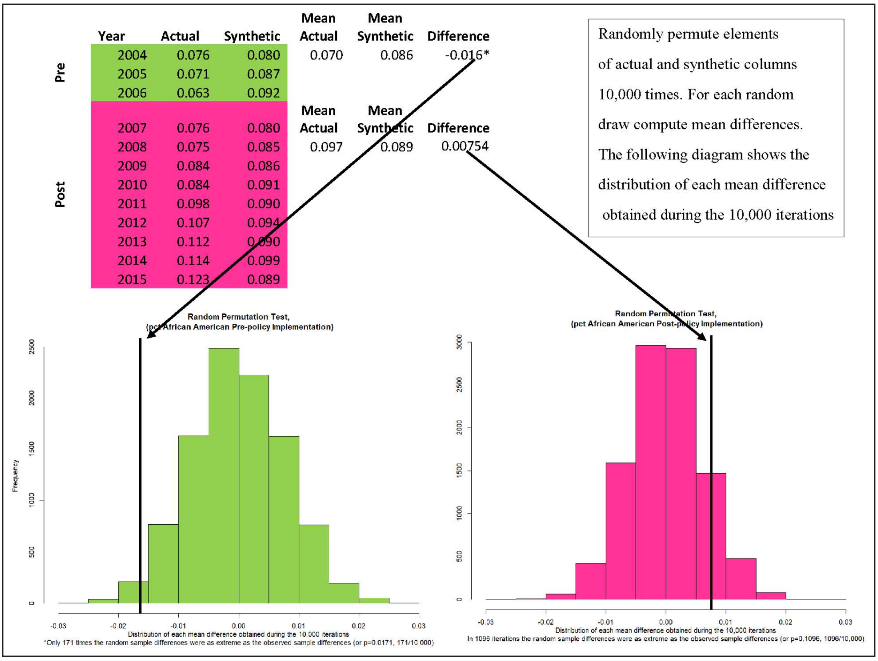


Figure 1. Rationale Behind the Implementation of Random Permutation Tests

mean differences. This suggests that before the implementation of the test optional policy, GMU had fewer Black students represented in their enrolled student population compared to its synthetic counterpart. The second histogram contains the analysis of the post-policy implementation outcomes. Note in this case that the random permutation test was conducted with 18 outcomes, nine of which corresponded to GMU and nine to its synthetic control during the 2007–2015 time period.[9] In this case, the negative difference found in the pre-implementation period became positive but did not reach statistical significance different from zero. This means that while the policy seemed to have helped increased representation of Black students, this difference is not strong enough. Indeed, there were almost 1,100 randomly generated differences that were more extreme than the observed difference found (0.0075). Nonetheless, note that Appendix A also includes a standardized mean differences test (Faraone, 2008)[10] to measure distances from the

[9]Due to data restrictions, Pell Grant and Average Institutional aid models (1 and 5) only cover the 2004 to 2014 time period, all remaining models cover 2004–2015.

[10]Following Faraone (2008) this procedure was implemented as follows: , wherein observed mean is the GMU pre- or post-implementation mean, and the standard deviation is obtained from all observations.

mean differences. This test indicated that the observed difference of 0.0075 resulting from these random permutation tests has an effect size of 0.58 standard deviations away from the mean. An effect size of this magnitude is an important indication of a meaningful increase in Black student enrollment after the test-optional policy implementation. Results for the permutation tests are included as Appendix A.

George Mason University's Synthetic Control Unit

The synthetic control unit for this study was derived from a donor pool comprised of institutions self-selected by GMU as their "custom comparison group" in their IPEDS Data Feedback Reports from 2005–2015. Nine of the 50 unique universities selected by GMU as its peers were private not-for-profit. Considering that these nine institutions have significantly dissimilar financial resources to GMU, including them in the creation of the synthetic controls would have increased heterogeneity that would be difficult to be modeled. Accordingly, following Abadie et al. (2010), we removed these nine institutions in order to maximize similarity between the donor pool and treated unit. We also removed seven public institutions from consideration for the donor pool due to missing data issues in both predictor and outcome variables (see Jaquette & Parra, 2014 for an analysis of related data issues). Ultimately, our final donor pool included 34 public institutions that required standardized testing for all undergraduate applicants prior to GMU's policy change.[11] Table 1 provides a list of these institutions and shows the distribution of weights given to each member of the donor pool. In a given model, the institution contributing the most to the synthetic GMU is bolded and will be utilized for the subsequent placebo test as explained in the "Placebo Tests" section above.[12]

Findings

This section contains the summary statistics for weighted and unweighted comparisons between GMU and its synthetic controls, as well as the results obtained from the SCM models.

[11]Some institutions in the donor pool adopted test-optional or test-flexible policies for admissions after GMU's policy was enacted (e.g., University of Kansas, University of Nevada–Las Vegas, and University of Nevada–Reno). These institutions can still contribute to the weighted control unit in the pre-treatment timespan since their data does not reflect this later policy intervention (Abadie et al., 2010).

[12]For Model 3, which considers Hispanic student enrollment, we used the second largest contributor (University of Connecticut) due to reciprocal issues when running the placebo model with the top-contributing institutions (University of Rhode Island).

TABLE 1.
WEIGHTS OF DONOR POOL INSTITUTIONS, BY DEPENDENT VARIABLE

Unit No.	*Comparison Institution*	*Weight*[1]	*Weight*[2]	*Weight*[3]	*Weight*[4]	*Weight*[5]
1	Florida State Univ.	0.000	0.323	0.000	0.271	0.000
3	Georgia State Univ.	0.000	0.000	0.107	0.000	0.000
4	Indiana Univ.-Bloomington	0.000	0.000	0.000	0.000	0.000
5	Ind. Univ.-Purdue Univ. Indianapolis	0.000	0.000	0.000	0.000	0.008
6	Michigan State Univ.	0.000	0.000	0.000	0.000	0.000
7	North Carolina State Univ.	0.000	0.000	0.000	0.000	0.000
8	Univ. of Arkansas	0.000	0.000	0.000	0.000	0.000
9	Univ. of Cincinnati	0.000	0.000	0.000	0.000	0.000
10	Univ. of Connecticut	0.010	0.000	0.141	0.000	0.311
11	Univ. of Florida	0.000	0.000	0.000	0.000	0.000
12	Univ. of Houston	0.000	0.000	0.000	0.000	0.000
13	Univ. of Illinois at Urbana-Champaign	0.000	0.000	0.000	0.000	0.000
14	Univ. of Iowa	0.000	0.000	0.000	0.000	0.000
15	Univ. of Kansas	0.000	0.175	0.000	0.187	0.000
16	Univ. of Kentucky	0.000	0.000	0.000	0.000	0.000
17	Univ. of Louisville	0.050	0.000	0.000	0.226	0.000
18	Univ. of Maryland	0.000	0.000	0.111	0.000	0.000
19	Univ. of Massachusetts Amherst	0.000	0.000	0.000	0.000	0.000
20	Univ. of Memphis	0.098	0.000	0.000	0.000	0.201
21	Univ. of Minnesota-Twin Cities	0.000	0.000	0.000	0.000	0.000
22	Univ. of Missouri-Kansas City	0.000	0.000	0.000	0.000	0.000
23	Univ. of Nebraska-Lincoln	0.000	0.000	0.000	0.000	0.000
24	Univ. of Nevada-Las Vegas	0.000	0.000	0.000	0.000	0.000
25	Univ. of Nevada-Reno	0.000	0.000	0.000	0.000	0.000
26	Univ. of North Carolina at Chapel Hill	0.156	0.000	0.000	0.063	0.000
27	Univ. of Oklahoma	0.000	0.000	0.000	0.000	0.000
28	Univ. of Rhode Island	0.686	0.000	0.641	0.233	0.478

Table 1. Cont.

Unit No.	*Comparison Institution*	*Weight*[1]	*Weight*[2]	*Weight*[3]	*Weight*[4]	*Weight*[5]
29	Univ. of South Carolina	0.000	0.000	0.000	0.000	0.000
30	Univ. of South Florida	0.000	0.000	0.000	0.012	0.000
31	Univ. of Tennessee-Knoxville	0.000	0.000	0.000	0.000	0.000
32	Univ. of Utah	0.000	0.000	0.000	0.000	0.000
33	Univ. of Wisconsin-Milwaukee	0.000	0.169	0.000	0.000	0.000
34	Wayne State Univ.	0.000	0.000	0.000	0.000	0.000
35	Western Michigan Univ.	0.000	0.333	0.000	0.000	0.000

NOTE: Bold value denotes highest contributing comparison institution in model, which is also used for placebo test
[1]Weighted based on %Pell students enrolled as dependent variable
[2]Weighted based on %Underrepresented minority students enrolled as dependent variable
[3]Weighted based on %Black students enrolled as dependent variable
[4]Weighted based on %Hispanic students enrolled as dependent variable
[5]Weighted based on Average Institutional Grant aid per FT student as dependent variable
[1]Weighted based on %Pell students enrolled as dependent variable
[2]Weighted based on %Underrepresented minority students enrolled as part of first-time FT cohort as dependent variable
[3]Weighted based on %Black students enrolled as part of first-time FT cohort as dependent variable

Summary of Weighted and Unweighted Statistics

In order to understand the comparability of the real GMU and its synthetic counterpart, Table 2 provides a breakdown of the predictor variable means for the treatment and control units in each model and the average of the 34 institutions in the donor pool. For each model, SCM weights the donor pool members to make the predictor and outcome variables within the synthetic unit as mathematically close as possible to the treated unit during the pre-treatment period. Notably, weighting of the donor pool can vary between models to ensure the synthetic control unit is the most mathematically similar to the treated unit as possible, which results in variation to the synthetic means.

The column titled "Real" in Table 2 shows the aggregate mean of each predictor from 2004 to 2006 for GMU, which accounts for the pre-policy time period observed. Similarly, the columns titled "Average of 34 Control Institutions" represents the aggregate average for all 34 donor institutions during the same pre-policy period. For example, GMU's average first year cohort was 2,416 students whereas the donor pool's average was 3,926 students. Note also that each of the five synthetic models aimed at reducing these observed difference in the unweighted means just described. More specifically, the mean difference between GMU and the unweighted average had a magnitude of approximately 1,520 students (or 3,926–2,416). This difference was consistently smaller in all but the second model, where it reached a magnitude of 1,753 (or 4,169–2,416). Although this magnitude increase in one model, this is not negative *per se* as the model will try to compensate for other predictors included. For example, the total price at GMU and the unweighted total is approximately $2,185 (or 16,907.8–14,722.7), but the gap between GMU and its synthetic control for model two is smaller at $1,202 (or 15,925.2–14,722.7). This pattern is found elsewhere in Table 2, as SCM will try to minimize the difference in the observed and weighted means of the synthetic controls (Abadie et. al., 2010).

SCM Findings

This section discusses the main findings of this study including five pairs of figures depicting changes in the outcome variables for treated and control units and placebo tests conducted to verify these findings. The first figure in each pair highlights the comparison between GMU and its synthetic version, with the vertical line denoting the 2007 policy implementation year. In each set of figures, the treated and control units were found to be statistically insignificant in the pre-treatment period using random permutation tests, which suggests that baseline equivalence was reached and that the post-treatment synthetic unit is a good approximation to a version of GMU that did not implement a test-flexible policy. Upon crossing the implementation

TABLE 2.
PREDICTOR MEANS, BY DEPENDENT VARIABLE

George Mason University Predictor Variables	*Real*	*Synthetic*[1]	*Synthetic*[2]	*Synthetic*[3]	*Synthetic*[4]	*Synthetic*[5]	*Average of 34 s Control Institution*
First Year Cohort	2416.000	2694.192	4648.291	2810.184	3895.418	2684.180	3926.039
Total Price	14722.667	17135.764	15187.513	18566.738	15609.989	17771.346	16907.814
State Appropriations per FTE	7214.417	9583.019	7575.123	9176.300	10113.954	9887.850	11037.405
Average SAT	1000.000	1010.173	998.051	1016.850	1023.662	1002.005	1018.284
%Admitted	0.661	0.688	0.761	0.667	0.690	0.668	0.710
%Yield	0.349	0.352	0.435	0.310	0.457	0.352	0.476
Average Inst. Grant per FY FT Student	6476.667	5254.118	2788.516	4917.646	4311.881	-	3867.716
%UR Minority Students	0.142	0.141	-	-	-	0.164	0.166
%Pell	0.205	-	0.218	0.206	0.210	-	0.232

[1]Weighted based on %Pell students enrolled as dependent variable
[2]Weighted based on %Underrepresented minority students enrolled as part of first-time FT cohort as dependent variable
[3]Weighted based on %Black students enrolled as part of first-time FT cohort as dependent variable
[4]Weighted based on %Hispanic students enrolled as part of first-time FT cohort as dependent variable
[5]Weighted based on Average institutional grant aid per FT student as dependent variable

period, it is expected the two lines will diverge and provide an inference on the effect of the policy. The second figure, which represents a placebo test, uses the member of the donor pool that contributed the most to the synthetic version of GMU as a treated unit and is compared to its synthetic counterpart. Outcome variation in the placebo tests should present no change before and after the policy implementation when compared to its placebo synthetic version. If the placebo test shows similar changes in magnitude and direction from the comparisons obtained between GMU and its synthetic control (original figure), the results of the GMU trend is deemed to be not the result of the policy change but the result of a different factor (Abadie et al., 2010). In effect, this invalidates inferences to be made about any effect resulting from the policy implementation. Finally, note that for each set of comparisons, the y-axis was normalized to have the same range. This was done for the purpose of improving readability and to ease comparison across GMU and its placebo estimates.

Pell enrollment. Figure 2a depicts trends in enrollment of Pell Grant recipients at GMU and its synthetic counterpart. Overall, Figure 2a shows that GMU's enrollment of Pell Grant recipients improved between 2007 and 2013 with a more pronounced increase after 2008 and reaching its peak in 2011. However, when compared to the variation observed in its synthetic version, GMU's increase in Pell enrollment was not statistically significantly different, therefore suggesting that the observed changes cannot be attributed to the implementation of the test-flexible policy. Instead, the increase in Pell enrollment was influenced by other factors that more uniformly impacted all institutions alike, which indicates GMU may have experienced a similar increase regardless of implementation of their test-optional policy. This statement is corroborated with Figure 2b, which uses the University of Rhode Island for the placebo test. In short, Figure 2b mirrors the trend and magnitude observed in Figure 2a, adding further uncertainty regarding the connection between the adoption of a test-optional admissions policy and Pell Grant enrollment.

Underrepresented minority enrollment. Figure 3a depicts trends in the percentage of students identified as part of an underrepresented minority group (Black, Hispanic, Native American, or Pacific Islander) that enrolled as part of GMU's and the donor institutions' first-time, degree-seeking, undergraduate cohort between 2004 and 2014. Unlike Figures 2a and 2b, the trajectory of the outcome variables in pre-treatment period in Figures 3a and 3b do not perfectly overlap. Despite this lack of perfect overlap, as noted above, random permutation tests still enabled corroboration of baseline equivalences in the pre-treatment period. Notably, immediately following policy implementation (in 2008), GMU exhibited a drop in the proportion of first-time full-time underrepresented students enrolled. Nonetheless, ever

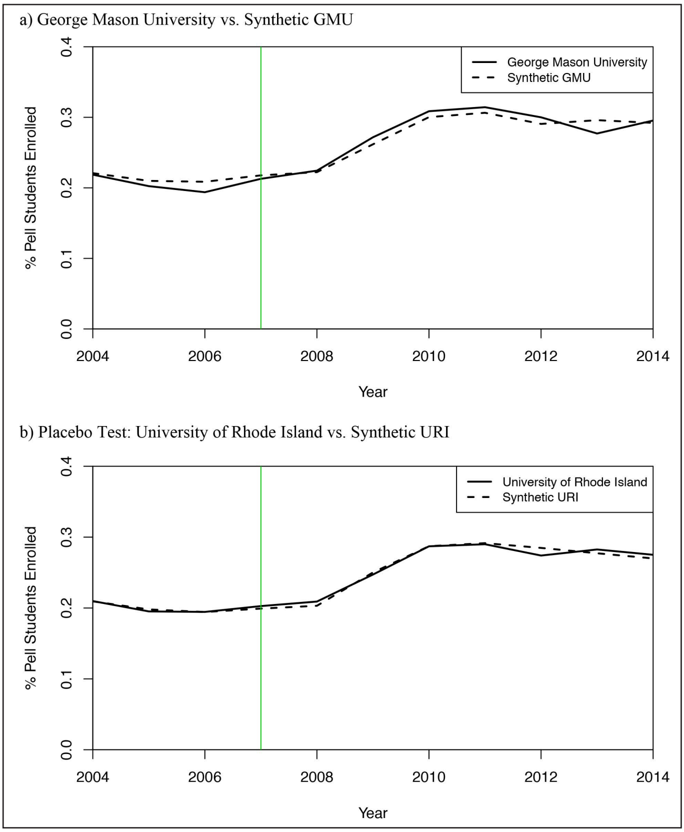


Figure 2. Trends in Percentage Pell Recipient Enrollment

since this drop, this proportion has followed an upward trajectory indicating an increase of the racial diversity of its student body. Once again, however, this upward trajectory was mirrored in the synthetic version. Figure 3b considered Western Michigan University for the placebo test. Note that the slope is considerably more positive than the slope observed for the synthetic WMU. This finding once again validates the notion that any observed increase for GMU's outcomes is not the result of the policy change. Despite this lack of evidence regarding the impact of the policy on student racial/ethnic diversity, it is worth considering that perhaps the policy impacted participants from some groups more others. Because of this possibility, we opted to disaggregate the dependent variable by focusing on the two main contributing groups–proportion of Black students and proportion of Hispanic students–to see if any statistically significant differences existed.

After disaggregating the dependent variable in Figure 3, Figure 4a depicts trends in the percentage of Black students among the first-time full-time

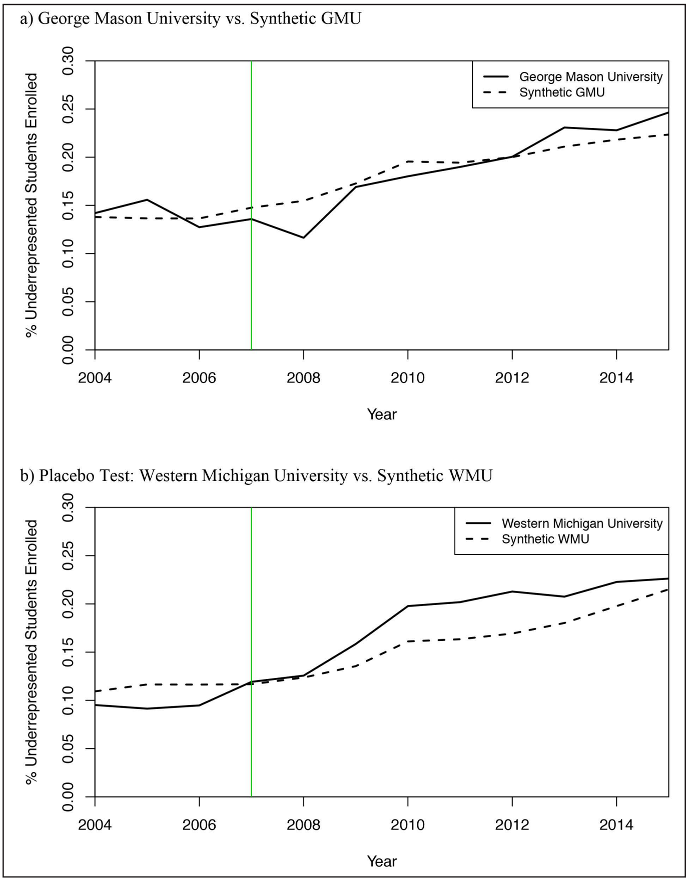


Figure 3. Trends in Percentage Underrepresented Student Enrollment

cohort enrolling at GMU between 2004 and 2015. Examination of differences in trajectories between Figures 3a and 4a justify the decision to disaggregate the analyses. Specifically, following the 2007 implementation year, the percentage of Black students enrolled has continued to improve at GMU, even when the synthetic version practically experienced no upward change during the entire period, including pre- and post-policy implementation. It is also worth noting that in the placebo test, wherein the University of Rhode Island served as the "treated" institution, both its trajectory and the trajectory of its synthetic version showed similar flat trends. While the finding shown in Figure 3a was not statistically significant, the improvement when compared to its synthetic version and its placebo estimate is noteworthy and will be discussed later.

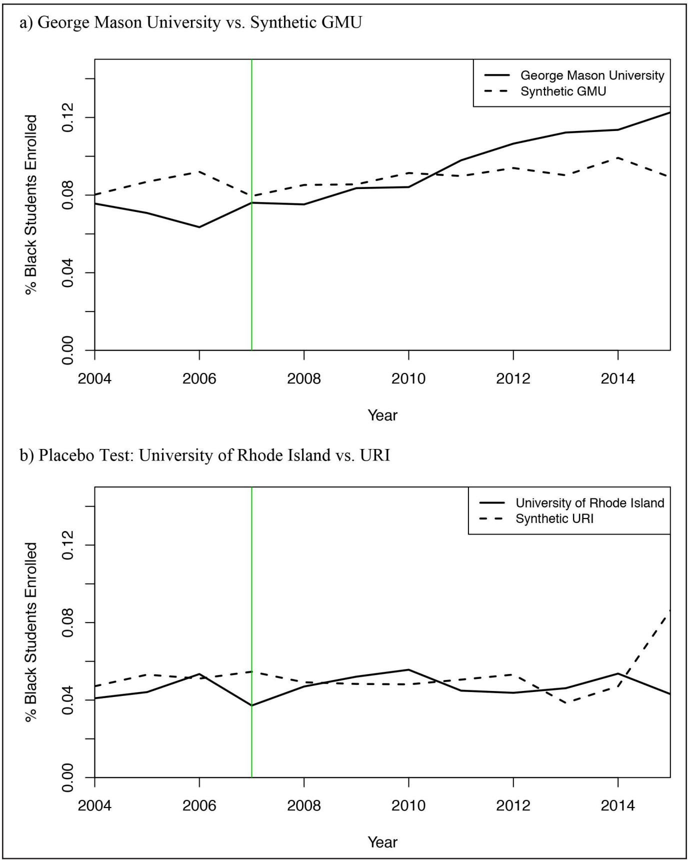


Figure 4. Trends in Percentage Black Student Enrollment

Turning to Figure 5a, GMU experienced a decrease in the percentage of Hispanic students in 2008, immediately after the implementation of the policy. Although there seems to be a modest increase since then, this increase was mirrored by its synthetic counterpart, therefore negating the possibility that such a change is due to the policy implementation. Figure 5b uses Florida State University for the placebo test. Note that the estimates consistently show a larger proportion of Hispanics in this institution compared to what its synthetic counterpart predicted. We argue that this difference in trajectories is merely a function of the location of Florida State University, which serves a larger proportion of Hispanic students. At any rate, the differences in the placebo test once again corroborate the lack of effect of the policy implemented at GMU to increase the representation of Hispanic/Latinx students.

Overall, following the adoption of a test-optional policy, GMU has seen a positive, but not statistically significantly different from zero, impact on the

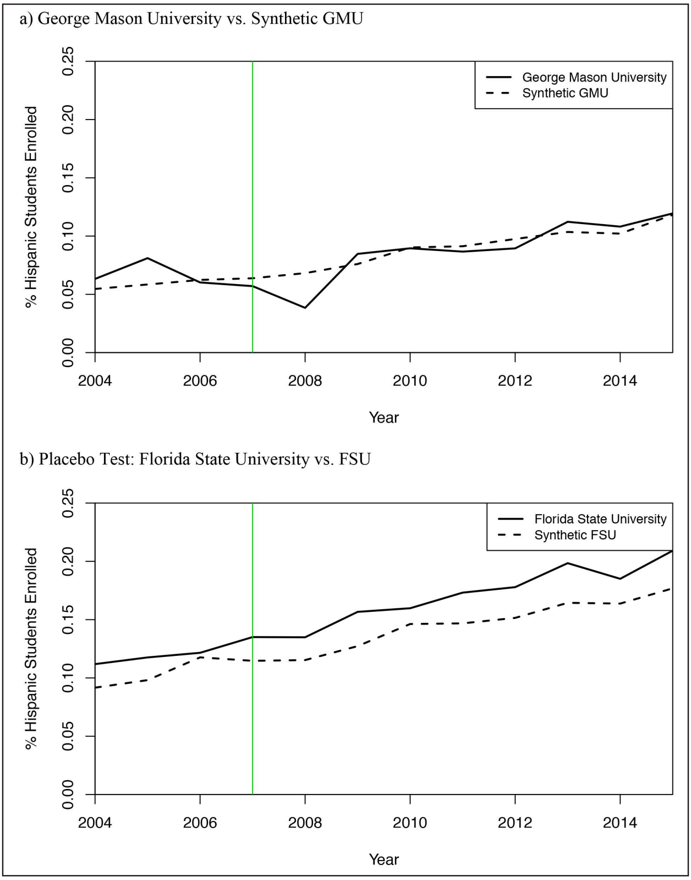


Figure 5. Trends in Percentage Hispanic Student Enrollment

enrollment of underrepresented students through 2014. While subpopulations, specifically Black students, have shown the greatest positive impact when compared to a synthetic version of the institution, representation of Hispanic students also improved albeit inconsistently. It is worth remarking that while these post-treatment changes are not considered to be the direct result of enactment of the test-optional policy, these trends in demographics are still important and may signal an indirect function of the adoption of the policy.

Institutional grant aid. The final model focuses on the average institutional grant per full-time student disbursed during the 2004 to 2014 time period. It is worth noting that this model, shown as Figures 6a and 6b, presents the highest variation in trends observed both pre- and post-policy implementation. GMU provided higher average grants per full-time student than the synthetic version predicted, but these average amounts decreased below the predicted amounts of its synthetic version in 2013. Although the gaps shown in Figure 6a may indicate a positive effect of the policy, note that a rather similar trend is shown in the placebo test. Once again, the placebo test enabled assessment of the validity of the trends found. In considering Figure 6b, which uses the University of Rhode Island for the placebo test, we conclude that the positive gaps are not the result of the implementation of the policy.

More specifically, there were differences in both the pre- and post-implementation periods in both Figure 6a and 6b. These results indicate that any changes observed in GMU cannot be attributed to the policy change but are the result of other factors and/or trends found in the data. In following with the falsification procedure upon which the placebo test is based, we found that in a setting wherein no policy was enacted, the fake site (University of Rhode Island in this case) was still doing better than its synthetic counterpart. This result provides more evidence against the potential claim that the policy was driving the results at GMU, given that an entity that did not implemented such a policy also had positive results, therefore invalidating any sort of causal link between test-flexible implementation and increase of average grant aid per FT student. Nevertheless, it remains noteworthy that GMU was doing better than its synthetic version both before and after the policy change, as depicted in Figure 6a, despite the policy not being the reason for this difference.

Although the finding from Figure 6 cannot be attributed to the policy change, it is still in contrast to what may have been expected by one of the guiding theories for this study, academic capitalism. Specifically, considering the minimal change in Pell recipient enrollment coupled with a positive, albeit statistically insignificant, increase in racial diversity, academic capitalism would predict GMU might be using their test-flexible policy to admit only students capable of paying sticker price. In turn, GMU could have saved institutional grant aid funding and create an additional revenue stream for other expenses. This scenario would predict a decrease in the average institutional grant per FTE student suggesting GMU has spent less institutional funds overall and per student since the policy change. However, as Figure 6a exhibits, there appears to be a minimal direct relationship between the policy and institutional aid.

LIMITATIONS

There are a variety of limitations to consider in regard to this study. First, the scope of students that applied through the test-flexible program at GMU is unknown. While this study aims to investigate the impact of the policy at

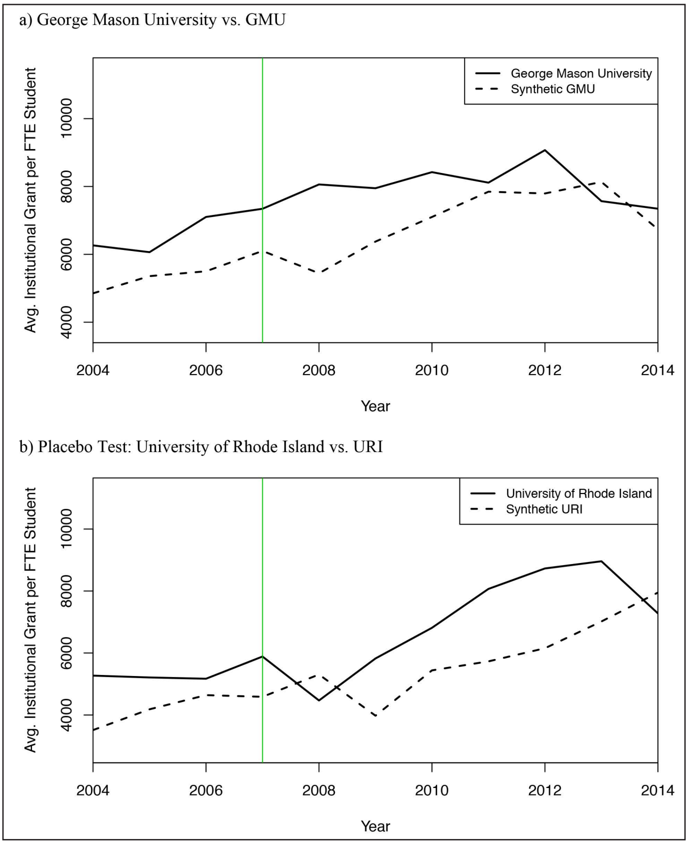


Figure 6. Trends in Average Institutional Grant per FT Student

the institution-wide level, an understanding of the extent to which this policy has impacted student applicant behavior could further our knowledge of the potential influence of test-flexible policies. That is, the dependent variables focused on enrollment, rather than application and/or acceptance, which could potentially contribute to the lack of statistical significance of these findings. In particular, test-optional and test-flexible policies may have a more direct relationship and greater influence on student application and/or admissions decisions rather than likelihood of enrollment, since the latter is based on an array of external factors and involvement of multiple parties. When possible, future research should focus on the impact of these types of policy on application behaviors. Lastly, due to data availability, there were only three years of data to examine pre-treatment variables. Although the synthetic control unit was statistically insignificant for each model, except for model 3, and therefore considered a good comparison overall, extending this timespan could have provided a greater perspective on trends prior to policy adoption.

Discussion

This study contributes additional insights to the larger dialogue around test-optional admissions policies, and provides potential avenues for future research. Overall, the findings for this study are in line with Belasco et al. (2014), regarding statistically insignificant changes to socioeconomic and racial diversity in student enrollment. While previously considered indicators of prestige and selectivity, such as applications received and *U.S. News & World Report* ranking, were not considered as a primary outcome, we found an inconclusive relationship between GMU's adoption of a test-flexible policy with variations in institutional grant aid, suggesting that this policy did not translate into disbursement of additional financial resources for the institution. In sum, although investigating a different institutional type and different version of test-optional policy (e.g., test-flexible), we found this policy to have minimal impact on enrolled student demographics. While this represents a common finding with other studies testing the effectiveness and influence of test-optional policies on increasing diversity, it may also highlight a broader finding regarding such goals.

Why Did This Policy Not Translate to Greater Diversity?

Considering the literature on postsecondary decision-making among traditionally underrepresented populations of students, our findings are not surprising. Returning to one of the guiding conceptual frameworks principles of this study, college access and choice theories, there are a plethora of additional considerations that deter underrepresented student populations from enrolling in college, including cost of attendance, knowledge and perception of the benefits of attaining postsecondary education, and potential geographic limitations. In fact, in discussing a policy adopted at the University of Virginia to increase low-income student representation, Tebbs and Turner (2006) highlighted three barriers to enrollment at public institutions: cost of attendance, information constraints, and long-term credit constraints (primary and secondary schooling and other resources that fail to promote collegiate attainment). Consequently, although test-optional and test-flexible policies eliminate specific barriers to applying to a given college, this policy alone will likely never result in the greater enrollment of traditionally underrepresented students into postsecondary institutions. In particular, financial obstacles and information constraints may be too severe for a test-optional or test-flexible policy alone to influence a student's postsecondary decision.

This limitation could be particularly notable in the case of a test-flexible policy like GMU's where high school GPA, class rank, and extracurricular involvement play a central role in qualifying to apply as a non-submitter. These criteria are typically ascribed to students who perform well on standardized tests and, therefore, may equate to a stronger commitment towards attracting traditional college-going students who did not attain "acceptable" standardized test scores, rather than improving college access. Alternatively, in line with academic capitalism, it remains possible that GMU's adoption of its test-flexible policy was motivated by the desire to attract academically strong students who could pay as close to the sticker price as possible, but might not have strong standardized test scores. This perspective is in line with the academic capitalism lens and market-like behaviors typically employed during recruitment events by colleges and universities (González Canché, 2017; Slaughter & Leslie, 1997; Slaughter & Rhoades, 2004). Future research on test-flexible policies should consider student characteristics of the pool of accepted applicants, prior to enrollment, to potentially examine the motivation guiding these policies. Unfortunately, while this question can be addressed with administrative data gathered by institutional research offices, IPEDS does not currently maintain this information on students throughout the application process.

Nevertheless, it remains notable that after GMU adopted its test-flexible policy, the institution increased their racial diversity in comparison to its synthetic version. Although the lack of statistical significance suggests these improvements were not directly the result of the policy, it is possible that the 2007 beginnings of test-flexible admissions signaled a shift in how GMU perceived the importance of racial diversity at the aggregate institution-level. Consequently, the adoption of a test-optional admissions policy may be an important factor driving an increase of diversity on campus, even if this was not GMU's originally stated goal. Unfortunately, while there has been work to further our understanding of what happens behind closed doors (Posselt, 2016; Steinberg, 2003), college admissions remain masked in a black box of decision-making by committees of admissions counselors. To this end, future work examining admissions personnel's perceptions and attitudes towards test-optional and test-flexible policies may indicate if these shifts in diversity are, in fact, less student-driven and actually a function of institutional decision-making.

Conclusion

Although proponents of test-optional and test-flexible policies argue that these policies can increase student diversity, in line with previous research, findings from this study suggest GMU's test-flexible policy has had a statistically insignificant influence in increasing the racial and socioeconomic diversity of its student body. This result is in accordance with the college access and choice models that guide this study. In particular, this literature notes standardized tests are only one of the many obstacles underrepresented populations must navigate prior to enrolling at a postsecondary institution. Consequently, recognizing the continued interest and adoption of policies deemphasizing standardized admissions testing, we close with recommendations for public college and universities that truly aspire to increase racial and socioeconomic diversity on campus.

First, test-flexible policies, such as the one assessed in this study, may not offer sufficient opportunity for underrepresented populations of students. In particular, due to the rigorous academic and non-academic criteria to qualify, these students might not know or be able to reasonably meet the necessary requirements based on opportunities available. Accordingly, public institutions that aim to increase diversity should instead consider test-optional policies, effectively removing the testing requirement for all applicants, or reduce the requirements to qualify for a test-flexible policy, such as relying on minimum GPA or high school ranking but not both. Nevertheless, an institution's adoption of a test-optional or test-flexible policy should not be viewed as a panacea to increase student diversity without considering other factors, such as academic preparation, geographic limitations, and financial obstacles. This implies that, while complete or conditional removal of testing requirements may improve the representation of underrepresented students in the applicant pool, the real challenge in closing higher education gaps for this population centers on ensuring that those applicants admitted are able to enroll and have the means necessary to persist and graduate. From this view, even complete removal of standardized admissions tests will not be sufficient to close the persistent gaps alone without the provision of adequate information, resources, and support for students to ensure their ability to thrive.

APPENDIX A.

RESULTS FROM RANDOM PERMUTATION TESTS (10,000 REPETITIONS)

	Comparison	*Pre-Policy Mean Difference (p-value*) [standardized mean differences] 2004–2006*	*Post-Policy Mean Difference (p-value) [standardized mean differences] 2007–2015*
Model 1[1**]	GMU vs. Synthetic GMU (Figure 1a)	-0.0081 (0.1519) [-0.88485]	0.0022 (0.4475) [0.062853]
	Univ. of Rhode Island vs. Synthetic URI (Figure 1b)	-0.0007 (0.4186)	0.0006 (0.4820)
Model 2[2]	GMU vs. Synthetic GMU (Figure 2a)	0.0047 (0.2518) [0.553635]	-0.00223 (0.4467) [-0.06706]
	Western Michigan Univ. vs. Synthetic WMU (Figure 2b)	-0.0202 (0.0084)	0.02327 (0.0914)
Model 3[3]	GMU vs. Synthetic GMU (Figure 3a)	-0.016 (0.0171) [-1.71639]	0.00754 (0.1096) [0.581595]
	Univ. of Rhode Island vs. Synthetic URI (Figure 3b)	-0.0042 (0.1295)	-0.0058 (0.1050)
Model 4[4]	GMU vs. Synthetic GMU (Figure 4a)	0.0096 (0.0736) [1.151136]	-0.00280 (0.3869) [-0.13335]
	Florida State Univ. vs. Synthetic FSU (Figure 4b)	0.0145 (0.0565)	0.0247 (0.0235)

Appendix A. Cont.

	Comparison	*Pre-Policy Mean Difference (p-value*) [standardized mean differences] 2004–2006*	*Post-Policy Mean Difference (p-value) [standardized mean differences] 2007–2015*
Model 5[5**]	GMU vs. Synthetic GMU (Figure 5a)	1239.891 (0.0153) [1.712392]	1042.027 (0.0103) [1.154176]
	Univ. of Rhode Island vs. Synthetic URI (Figure 5b)	-1104.408 (0.0188)	1232.884 (0.0467)

*p-value is the proportion of times that the difference between the treated and control units had a greater magnitude than the actual observed difference

**Due to data restrictions, these models only cover the 2004 to 2014 time period, all remaining models cover 2004-2015. Standardized mean differences were conducted as follows: (actual mean–synthetic mean)/pooled standard deviation

[1] Model 1: %Pell students enrolled as dependent variable

[2] Model 2: %Underrepresented minority students enrolled as part of first-time FT cohort as dependent variable

[3] Model 3: %Black students enrolled as part of first-time FT cohort as dependent variable

[4] Model 4: %Hispanic students enrolled as part of first-time FT cohort as dependent variable

[5] Model 5: Average institutional grant aid per FT student as dependent variable

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 492–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2011). Synth: An R package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, *42*(13). Available at SSRN: *http://ssrn.com/abstract=1958891*.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, *93*(1), 112–132.

Atkinson, R. C. (2001, February 18). *Standardized tests and access to American universities*. The 2001 Robert H. Atwell Distinguished Lecture, delivered at the 83rd Annual Meeting of the American Council on Education. Washington, DC. Retrieved from: *http://www.ucop.edu/news/sat/speech.html*.

Balf, T. (2014, March 6). The story behind the SAT overhaul. *New York Times Magazine*. Retrieved from *http://www.nytimes.com/2014/03/09/magazine/the-story-behind-the-sat-overhaul.html*.

Bates College. (2004, October 10). *PowerPoint analysis: SAT submitters and non-submitters*. Retrieved from *http://www.bates.edu/news/2004/10/10/powerpoint-analysis/*.

Belasco, A. S., Rosinger, K. O., & Hearn, J. C. (2014). The test-optional movement at America's selective liberal arts colleges: A boon for equity or something else? *Educational Evaluation and Policy Analysis*, *20*(10), 1–18.

Blau, J. R., Moller, S., & Jones, L. V. (2004). Why test? Talent loss and enrollment loss. *Social Science Research*, *33*, 409–434

Breneman, D. W., Doti, J. L., & Lapovsky, L. (2001). Financing private colleges and universities: The role of tuition discounting. In M. A. Paulson & J. C. Smart (Eds.), *The finance of higher education: Theory, research, policy, and practice* (pp. 461–479). New York, NY: Agathon Press.

Camara, W. J. & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification* (College Board Report No. 1999–5). New York, NY: The College Board.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (Vol. 2). Pacific Grove, CA: Duxbury.

Crouse, J. & Trusheim, D. (1988). *The case against the SAT*. Chicago, IL: The University of Chicago Press.

Ehrenberg, R. (2002). Reaching for the brass ring: The U.S. News & World Report rankings and competition. *The Review of Higher Education*, *26*, 145–162.

Epstein, J. (2009). Behind the SAT-optional movement: Context and controversy. *Journal of College Admission*, 9–19.

Faraone, S. V. (2008). Interpreting estimates of treatment effects: Implications for managed care. *Pharmacy and Therapeutics*, *33*(12), 700–711.

Fleming, J. (2002). Who will succeed in college? When the SAT predicts black students' performance. *The Review of Higher Education*, *25*(3), 281–296.

Freedle, R. O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, *73*, 1–43.

Garvey, J. E. (1981). *Testing for college admissions: Trends and issues*. Arlington, VA: Educational Research Service, Inc.

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, *8*(1), 1–26.

George Mason University. (2006, May 25). George Mason adopts SAT-optional admissions policy. *The Mason Gazette*. Retrieved from *https://gazette.gmu.edu/articles/8464*.

George Mason University. (2017). Freshmen admissions. Retrieved from *https://www2.gmu.edu/admissions-aid/how-apply/freshman*.

González Canché, M. S. (2014). Localized competition in the non-resident student market. *Economics of Education Review*, *43*, 21–35. *http://dx.doi.org/10.1016/j.econedurev.2014.09.001*.

González Canché, M. S. (2017). The heterogeneous non-resident student body: Measuring the effect of out-of-state students' home-state wealth on tuition and fee price variations. *Research in Higher Education*. *http://link.springer.com/article/10.1007/s11162–016–9422–2*.

González Canché, M. S. (2019). Geographical, Statistical, and Qualitative Network Analysis: A Multifaceted Method-Bridging Tool to Reveal and Model Meaningful Structures in Education Research. In Paulsen, Michael B. & Perna, Laura (Eds), *Higher Education: Handbook of Theory and Research*. vol 34. Springer, Cham.

Good, P. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses, second edition*. New York, NY: Springer.

Hearn, J. C. (2001). Access to postsecondary education: Financing equity in an evolving context. In M. B. Paulsen & J. C. Smart (Eds.), *The finance of higher education: Theory, research, policy, and practice* (pp. 439–460). New York, NY: Agathon Press.

Hearn, J. C. (2006). Enhancing institutional revenues: Constraints, possibilities and the question of values. In R. L. Clark & M. B. d'Ambrosio (Eds.), *The new balancing act in the business of higher education* (pp. 27–45). Northampton, MA: Edward Elgar Publishing.

Hiss, W. C., & Franks, V. W. (2014). *Defining promise: Optional standardized testing policies in American college and universities admissions*. Arlington, VA: National Association for College Admission Counseling. Retrieved from: *http://www.nacacnet.org/research/research-data/nacac-research/Documents/DefiningPromise.pdf*.

Hoffman, J. L., & Lowitzki, K. E. (2005). Predicting college success with high school grades and test scores: Limitations for minority students. *The Review of Higher Education*, *28*(4), 455–474.

Hossler, D., & Gallagher, K. S. (1987). Studying college choice: A three-phase model and the implications for policy-makers. *College and University*, *62*(3), 207–221.

Jaquette, O., & Parra, E. E. (2014). Using IPEDS for panel analyses: Core concepts, data challenges, and empirical applications. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 29, pp. 467–533). Netherlands: Springer.

Jencks, C. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap*. Washington, DC: The Brookings Institution Press, 55–85.

Kapor, M., & Klein, F. K. (2007, November 4). Early admissions policies give children of the rich an edge. *The San Francisco Chronicle*. Retrieved from *http://www.sfgate.com/opinion/article/Early-admissions-policies-give-children-of-the-3234970.php*.

Klasik, D. (2013). The ACT of enrollment: The college enrollment effects of required college entrance exam testing. *Educational Researcher*, *42*(3), 151–160.

Manski, C. F., & Wise, D. (1983). *College choice in America*. Cambridge, MA: Harvard University Press.

Phipson, B., & Smyth, G. K. (2010). Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, *9*(1), Article 39.

Posselt, J. R. (2016). *Inside graduate admissions: Merit, diversity, and faculty gatekeeping*. Cambridge, MA: Harvard University Press.

Robinson, M. & Monks, J. (2005). Making SAT scores optional in selective college admissions: A case study. *Economics of Education Review*, *24*, 393–405.

Rosner, J. (2012). The SAT: Quantifying the unfairness behind the bubbles. In J. Soares (Ed.), *SAT wars: The case for test-optional college admissions*. (pp. 104–117). New York, NY: Teachers College Press.

Slaughter, S., & Cantwell, B. (2012). Transatlantic moves to the market: The United States and the European Union. *Higher Education*, *63*(5), 583–603.

Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university.* Baltimore, MD: Johns Hopkins University Press.

Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy: Markets, state, and higher education*. Baltimore, MD: Johns Hopkins University Press.

Soares, J. A. (2012). For tests that are predictively powerful and without social prejudice. *Research and Practice in Assessment*, *7*, 5–11.

Steinberg, J. (2003). *The gatekeepers: Inside the admissions process of a premier college*. New York, NY: Penguin Books.

Syverson, S. (2007). The role of standardized tests in college admissions: Test-optional admissions. *New Directions for Student Services*, *118*, 55–70.

Tebbs, J., & Turner, S. (2006). The challenge of improving the representation of low-income students at flagship universities. In M. McPherson & M.O. Schapiro (Eds.), *College access: Opportunity or privilege?* (pp. 103–115). New York, NY: The College Board.

Zemsky, R., & Oedel, P. (1983). *The structure of college choice*. New York, NY: College Entrance Examination Board.

Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education*. New York, NY: RoutledgeFalmer.

Zwick, R. (2004). Is the SAT a "Wealth Test"? The link between educational achievement and socioeconomic status. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 203–216). New York, NY: RoutledgeFalmer.

Zwick, R., & Green, J. G. (2007). New perspectives on the correlation of scholastic assessment test scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement*, *44*(1), 1–23.

2018

DEFINING ACCESS:

How Test-Optional Works

Co-Authors: Steven T. Syverson, Valerie W. Franks, William C. Hiss

To be hosted electronically by the National Association for College Admission Counseling

AUTHORS' NOTE

This is a continuation into the investigation on test-optional policies, started by conducting individual case studies on 28 public and private colleges and universities in the United States. In every instance that we have presented an observation or comparison that sheds light on the use of test-optional admission policies, we have included every institution that provided reliable data for the particular comparison.

But unlike a study which captures a single database and conducts a series of analyses on that data, we have collected data from institutions that became test-optional more than a decade and a half ago, and others that adopted test-optional policies within the last few years. As a result, only a minority of our analyses draw on the full 28 institutions. We have tried to provide as wide a framework of findings as possible, while identifying for each analysis the number of institutions and student records that were included.

Please look for the explanation in the figure description of each chart on how that subset of institutions was selected. Though we've provided connecting narrative, this report can best be considered an anthology of short reports designed to provide insights into the use of test-optional college admissions policies in the nation in the past decade.

Though the participating institutions may choose to remain anonymous, we wish to publicly thank the deans of admission and particularly the Institutional Research staff at each of these colleges for their extraordinary commitments in helping us to assemble and interpret this massive amount of data in an effort to better understand how test-optional policies are working at their institutions.

"This Commission wishes to emphasize at the outset that a "one-size-fits-all" approach for the use of standardized tests in undergraduate admission does not reflect the realities facing our nation's many and varied colleges and universities. These institutions differ greatly in size, selectivity and mission. At some, standardized tests are important predictors of students' academic success, while at others, they add little compared to high school grades."

NACAC Commission on the Use of Standardized Tests in Undergraduate Admission, 2008

EXECUTIVE SUMMARY

The number of colleges using **Test Optional Policies (TOPs)** in higher education admissions has dramatically expanded in recent years. And these colleges have avoided "one-size-fits-all," finding varied ways to administer TOPs and experiencing varied outcomes. Much of the momentum around Test-Optional admission is focused on whether the use of standardized tests (specifically SAT and ACT) **unnecessarily truncates the admission of otherwise well-qualified students**. In particular, there is concern about whether widespread reliance on the use of these tests in the admission process tends to replicate the status quo in social class and opportunity in our American society.

In this study, we collected student-record level data from 28 institutions that illustrate the variety among institutions that have adopted a TOP. They ranged in undergraduate enrollments from 1,500 to 20,000 and 15%-90% in admission selectivity, and included long-time users of TOP as well as recent adopters of the policy. In most instances we received four cohorts of student data, in total representing a dataset of 955,774 individual applicant records. We focused on interpreting the data using practical significance rather than experimental statistical techniques.

A TOP was described by many of the admission deans of the participating institutions as a tool they employed in the hope of increasing applications from a more diverse range of students, so this report focuses great attention on traditionally under-represented populations in American higher education. To do so, we used our record-level data to identify the **intersectionality of these underserved populations**: First-Generation College Bound, students from lower SES backgrounds (Pell recipients as proxy), and students from racial and ethnic groups that have traditionally been underrepresented in college populations (URM). We identified students associated with *any* of these three groups and designated them as a single category of "Expanded Diversity," and when possible, used it in our explorations.

The experiences of institutions in this study provide evidence that the adoption of a well-executed test-optional admission policy can lead to an ***increase in overall applications*** as well as an ***increase in the representation of URM students*** (both numeric and proportionate) in the applicant pool and the freshman class. Roughly two-thirds of our TOP institutions experienced URM growth above that of a matched test-requiring peer institution. A similar but smaller magnitude increase was seen among Pell recipients.

Approximately **one quarter of the students in this study did not submit standardized test scores** with their college application (henceforth to be referred to as "Non-Submitters"). As noted in earlier studies, **URM, First-Generation-to-**

College, and **Pell recipients** were more strongly represented among Non-Submitters. For instance, **35%** of Black or African-American students chose to be Non-Submitters (12 percentage points higher than the overall non-submitting rate), as compared to **18**% of white students. Similarly, women chose to be Non-Submitters at higher rates than men.

We also found that Non-Submitters were often **admitted at *lower* rates** than Submitters, but, on average, **enrolled (yielded) at substantially *higher* rates**. Their HSGPAs were modestly lower than the Submitters, and, upon entering college, their First Year GPAs and Cumulative GPAs were comparably lower. However, **they ultimately graduated at rates equivalent to, or marginally higher than, Submitters**, the ultimate proof of success.

Furthermore, our data indicated that **high school GPA had a stronger correlation with college success for Non-Submitters than the ACT/SAT** (for the 27% of Non-Submitters for whom we had test scores) -- both in terms of college cumulative GPA and graduation rate. While test scores had a generally stronger relationship with college GPAs for the Submitters, for the Non-Submitters they tended to show a weaker relationship, essentially under-predicting the college GPA. The test scores continued to most strongly correlate with family income.

A financial analysis, though inconclusive, suggested that some degree of financial investment was required to support the success of a TOP policy. While the proportion of students with need did not necessarily increase after policy adoption, average demonstrated need and gift aid per capita did. **Non-Submitters were generally needier** than Submitters. They also, however, included a **sizable proportion of No-Need students,** only modestly lower than that of Submitters. We noted that well over half of all No-Need students were offered some gift aid, but No-Need **Non-Submitters were less likely than Submitters to receive gift awards**, in spite of the fact that these two groups were shown to graduate at comparable rates.

We cannot lay claim to definitive conclusions about the workings of a test-optional admission policy. However, our findings suggest that a **TOP works well for many types of institutions**. It appears to offer a less obstructed path to higher education for this population of students who feel that their scores do not match their ability. We do not argue that institutions should entirely eliminate consideration of the ACT and SAT for all their students, however, we do continue to **question whether the value-add of testing is large enough to justify the price—time spent, financial cost, and emotional drain—being paid by students** due to societal preoccupation with these tests.

We find that there is plenty of room in American higher education for diversity of process, allowing test-optional admission to contribute to that diversity. Some have asked, 'Why ignore a piece of information if it is available?" And we agree. Indeed, when a student chooses to be a Non-Submitter, that, too, is a valuable piece of information.

> *"We have great pride drawn from how well TOP works for first gen and diversity, and kids with special talents. For us, removing the testing was a big help in looking at everything else."*
>
> *Dean from a small private college*

OBSERVATIONS ON THE TEST-OPTIONAL MOVEMENT

A rapidly increasing number of colleges and universities have adopted test-optional admission policies, or TOPs, that allow some or all of their applicants to refrain from submitting standardized test scores. The institutions that officially deemphasize standardized tests in admission now total more than 1,000, including over 100 more not-for-profit institutions[1] in the past four years (FairTest List, 2018). From early adopter TOP institutions (Bowdoin in 1969, and Bates in 1984) to those recently adopting a TOP (Wesleyan and Bryn Mawr, both in 2014, George Washington University in 2016), some of the institutions choosing a TOP have national reputations for excellence. But TOP is not used just by highly selective private institutions—the FairTest list covers a range of public, regional private, and also for-profit institutions. A wide variety of institutions have found TOPs to be workable, productive tools to support their enrollment planning.

The momentum of the Test-Optional movement appears to be fed in part by several overlapping changes in how academic promise and outcomes are being evaluated. Collectively these changes are moving admission decisions away from heavy reliance on measures increasingly deemed to provide a narrow assessment of human potential. Many K-12 schools are moving toward proficiency and standards-based evaluations. Thousands of high schools have stopped providing Class Rank, as a false or misleading measure. Many colleges and universities are broadly moving to "holistic" admissions philosophies, using TOPs, and versions of "portfolio" admissions with new technologies like the Master Transcript that encourage students to provide evidence of individual talents and commitments. Collectively, these changes are decreasing the reliance on test scores and class rank to guide college admission decisions and guidebook rankings. Experiments are drawing on the findings of Admissions reform groups like the Institute on Character and Admission, or several on-going research projects: the "Turning the Tide" Project at Harvard, the Character Skills Snapshot from the Enrollment Planning Consortium, or the Master Transcript Consortium.

As this policy has become more pervasive, researchers have investigated its relative success. A 2015 study out of the University of Georgia (Belasco, 2014) found that, at the aggregate level, selective liberal arts colleges that adopted a TOP (in comparison with those that continued to require SAT or ACT scores) had <u>not</u> increased their enrollment of URM students or Pell recipients. But in focusing on these high-level, averaged outcomes, that study may not have been able to discern impact at the institutional level.

There has also been a proliferation of research on standardized testing in admission focused on the predictive value of testing and its fairness relative to various subpopulations of students. Much of the research affirming the value of the SAT and ACT has been conducted by the testing organizations. One such study, a synthesis of recent test score validity studies (Mattern and Patterson, 2014), states that the SAT provides incremental validity above and beyond HSGPA in the prediction of cumulative GPA, retention and graduation. Their conclusion: the combination of HSGPA and testing will produce the most accurate predictions of college success. Another recent release, a just-published volume of essays, *Measuring Success: Testing, Grades and the Future of College Admissions* (Buckley, Letukas, and Wildavsky, 2018) is principally a response by the College Board and ACT to the rapid growth of TOPs.

[1] The FairTest list includes many colleges that are "For Profit," and others that are "Test Flexible" -- allowing applicants to choose which form of testing to submit. The 28 institutions in this study, all Not-for-Profit and none using a "Test Flexible" policy, are drawn from two groups on the FairTest lists: the 129 National Liberal Arts Colleges and National Universities, and the 174 Regional Colleges and Universities.

Challenges to the pervasive use of these tests, their actual value, and their negative impact on students have come in a number of books (*Crossing the Finish Line: Completing College at America's Public Universities (*Bowen, Chingos, McPherson, 2009)*; SAT Wars: The Case for Test Optional Admissions (*Soares, 2012), the work and recommendations of the NACAC Commission on the Use of Standardized Testing in Undergraduate Admissions, the ongoing work of FairTest, a thoughtful documentary film released in early 2018, "The Test and the Art of Thinking" (Davis, 2018), and a variety of articles and smaller research projects.

In particular, the exhaustive research available in <u>Crossing the Finish Line</u> has been centrally important in this discussion. The authors, the late William Bowen and Michael McPherson, the former Presidents of Princeton and Macalester, respectively, with their research colleague Matthew Chingos, addressed what characteristics predicted graduation at a group of large public universities. Their data was drawn from institutions that required standardized tests from all students; none of the universities they studied had adopted a "threshold" admissions policy, with automatic admission granted to students who meet cut-off requirements for HSGPA. In the chapter which examined the predictive value of test scores and high school grades, they reported that:

> *The findings are dramatic. …the coefficients for SAT/ACT scores are always less than 0.02, which means that an increase in test scores of one standard deviation is associated with an increase of less than 2 percentage points in six-year graduation rates; this relationship is even negative at the historically black colleges and universities (HBCU's)…. The consistency of the results is extraordinary: In all but one of these more than 50 public universities, high school GPA remains a highly significant predictor of six-year graduation rates after taking account of the effects of test scores... Test scores, on the other hand, routinely fail to pass standard tests of statistical significance when included with high school GPA in regressions predicting graduation rates, especially when we leave the realm of the most highly selective public universities… …the remaining incremental predictive power of the SAT/ACT scores disappears entirely when we add controls for the high school attended, whereas the predictive value of the high school GPA increases.* (Bowen, Chingos, McPherson, 2009)

In 2014, William Hiss and Valerie Franks —two of the three co-authors of this study—released *Defining Promise: Optional Standardized Testing Policies in American College and University Admission* (Hiss, Franks, 2014). It was the first large-scale, multi-institution assessment of the outcomes of optional testing, and extended the research done in 2010 by Hiss and his co-author Kate Doria, in a 25-year look-back study on the outcomes of the policy at Bates College (Hiss, Doria, 2010).

The 2014 research revealed that—when given the option at one of those 33 TOP institutions —roughly a third of enrolled students chose to apply without standardized test scores (Non-Submitters). These Non-Submitters went on to graduate at virtually the same rates (a 0.6% difference) and with nearly the same college GPA (0.05 of a Cum GPA point) as the Submitters whose test scores were considered in the admission process. Their research also concluded that Non-Submitters were more likely to be first-generation-to-college, underrepresented minority students, women, Pell Grant recipients, and students with Learning Differences. And, using large volumes of HSGPA data, their findings underscored the sturdiness of the HSGPA as a predictor of college performance.

This research highlighted an interesting intersection between the testing agencies and that of the counter views. A meta-analysis of studies of "discrepant performance" revealed that "a quarter to a third of tested students exhibit some degree of mismatch between their grades and their test scores." Within this group, approximately half of them have high school grades that are higher than what the tests would predict. Across the studies cited, the range appears to be between 11% and 18% of the sample population (Sanchez & Mattern [Making the Case for Standardized Testing] in Buckley, 2018).

Another related study identified the students most likely to have strong HSGPAs and low testing: women, First-Generation to college, low income students, and students who speak a second language at home. Furthermore, those most likely to be discordant with weaker HSGPAs and stronger testing are males, whites, and those of higher income (Sanchez and Edmunds, 2015).

We would emphasize that the results only include those students who took the tests. It is quite plausible that there are other students who might have succeeded in college, had they been encouraged, found the right mentor, or were not so discouraged by the costs. If so, the real "discrepancies" may be even higher.

And it is worth noting the parallels between the proportions noted in the above studies and the median percentage (23%) of students choosing to be Non-Submitters at the institutions in this study. Does TOP indeed serve this "discrepant" population of students? Does it reduce admission barriers for underserved populations? The following study design continues with the full list of research questions and explains how we went about answering them.

STUDY DESIGN

We began this study by reviewing the FairTest list of 1000 colleges and universites with one form or another of optional testing. In contrast to our 2014 study, we eliminated any highly-specialized institutions (e.g., art institutes). We tried to achieve a balance of institutional types, sizes, selectivity, and geography, as well as to have representation from institutions with different approaches to TOP. We approached about one hundred institutions to discuss whether they would consider joining the study. We launched the study with 28 interested institutions that felt they could provide reliable data for the multiple cohorts needed for the study.

To give us context, we interviewed the Dean of Admissions or Enrollment VP at each institution in the study about their rationale for adopting a test-optional admission policy as well as commentary about how well they felt it was working at their institutions. In many instances, the current dean had not been at the institution at the time of adoption, so they relayed their best understanding of the circumstances at the time of adoption.

We received a large set of data: from the 28 colleges and universities, we received 955,744 student records, with up to 40 data items in each student record. With any dataset of this size, there will be elements that require careful examination and decisions about clarity and reliability. However, all data in this study comes from participating colleges and universities or from IPEDS. All data was carefully checked and cleaned for consistency and accuracy, leading in almost every case to clarifying follow-up communications with Institutional Research or Admissions research directors. In some instances, new coding schema or changes in computer systems interfered with the institution's abililty to provide comprehensive information across the span of the study.

We use commonly accepted statistical methologies (descriptive statistics, t-tests, chi-squares, Cohen's d) to present data and highlight statistical significance, but we have avoided highly complex "semi-experimental" statistical methodologies. Rather, we try to present the data in a straightforward fashion: Which students were drawn to being Non-Submitters of testing? How did Submitters and Non-Submitters compare in high school achievement, and subsequently in college performance? Did adoption of a TOP have an impact on the institution's applicant pool or enrolled classes? Did these institutions treat Submitters and Non-Submitters differently in their admission and aid policies?

The study was not designed to come to a single conclusion about the use of test-optional admission policies, but to explore as many dimensions around the policy as possible. We began by conducting individual case-study analyses for the 28 public and private colleges and universities. They have all been guaranteed anonymity, so this report uses aggregated data from subsets of institutions and avoids institutionally-identifiable information. Below each chart or diagram is a description of the number of institutions and records included, along with a brief profile. In every case we have included all the institutions that had reliable data for the analysis being presented. This report is a series of observations, rather than a series of parallel findings on a single set of data.

What are the principal research questions?

In an effort to shed additional light on the impact of a test-optional admission policy, this report explores several pertinent questions about test-optional admission:

- If an institution adopts a test-optional admission policy, does it reduce admission barriers, thereby encouraging more students to apply?
- Does adopting a test-optional admission policy help an institution attract and enroll more traditionally under-represented minority (URM), first-generation-to-college, and low-SES students?
- How do institutions "treat" students who have chosen to withhold their scores from the admission process – in both their admission decisions and their aid decisions?
- Are there institutional financial implications to adopting a test-optional admission policy?
- Who are the Non-Submitters -- the students who use a test-optional policy? How do they perform academically in college compared to students who do not use the policy? This portion of the study is a re-testing of the findings from "Defining Promise", with a largely different group of institutions, but a parallel methodology.

We are conscious of, and accept, the responsibility to have this study examine both the ethical issues like access and diversity, and the strategic issues of yields, changing classes and potential financial impact on the institution.

What types of institutions and policies are represented in this study?

We focused our participant recruitment on 4-year, degree-granting, IPEDS-submitting, public and private not-for-profit institutions in the United States. We then investigated the breadth of test-optional (TOP) policy types employed by institutions in the U.S. There is no standard definition of "test-optional admission," leaving institutions to define and implement a variety of policies. We organized the various versions of the policy in common categories, and we found, in approximate numbers, the primary types of TOP used by institutions (in rank order of frequency they were observed): *Academic Threshold, Optional for All, Optional Plus, Optional for Some,* and *Test Flexible*. The most commonly used policies—Optional for All and Academic Threshold—were of particular interest, as was the Optional Plus policy. Institutions with Test Blind and Test Flexible policies were not considered for inclusion in this study. While considered, no Optional for Some institutions were included in the study due to small numbers of Non-Submitters. Figure 1, below, describes each policy in more detail, estimates the proportion of that policy type represented in the U.S., and then counts those represented in the present study.

TEST OPTIONAL POLICY TYPES

Optional for All Policy	Optional Plus Policy	Optional For Some Policy	Academic Threshold Policy	Test Flexible Policy	Test Blind
				Not Qualified for Participation in Study	
• Allows most or all students to choose whether or not to submit standardized testing scores as part of their admissions application. - If any exceptions are made, homeschool and international students are required to submit some form of test (SAT, TOEFL, ACCUPLACER, etc.).	• Non-submitters are required to supplement their application with an interview or extra writing samples. - Examples of replacement submissions include: Interviews, non-cognitive tests, extra essays, graded papers.	• Testing options are offered to some student groups, but not others. For example, the policy is available to students unless… - Student wishes to be considered for scholarships / aid - student wishes to enroll in a particular school (nursing, engineering, etc.) - student is from out of state.	• Students who meet certain academic criteria (e.g., rank, GPA) are admitted without standardized testing scores as part of the admissions decision. - Often, but not exclusively, used by public institutions, and called Assured or Guaranteed Admission.	• Students have the option to submit scores from other testing in place of the SAT or ACT. - Examples of replacement tests include: International Baccalaureate, Advanced Placement, SAT II, Regents.	• Scores may be submitted, but they will not be reviewed by admission staff to make the admission decision. - Rarely used, mostly by small private institutions
Approximate Proportions on FairTest List					
39%	9%	5%	46%	1%	<1%
# Represented in the Study					
17	9	0	2	0	0

Figure 1. Test-Optional Policy Types, as of Spring 2016
This investigation and categorization focused on IPEDS Reporting, Degree granting, 4-year Public and Private Not-for-Profit Institutions on the FairTest List.

Some institutions (including many public universities) have a required HSGPA or class rank for Non-Submitters, often a requirement from a state education agency or the legislature. We have referred to this form of TOP as "academic threshold," in that the students must meet the required HSGPA for automatic admission. Shaped largely by the ability of particular threshold institutions to provide data, our research in the 2014 study had more of these "threshold" public institutions than this 2018 study. The best known of these are the Texas "Top 10%" policy at all Texas public universities, and the GPA requirement at the California State University system. Those requirements often become a self-fulfilling prophecy of student success in college. While sometimes contentious, a number of these policies have been in place for many years and seem to work successfully to open these public universities to a wider pool of applicants.

It should be noted that some institutions could fit the definitions of a couple of categories and/or have migrated from one to another. For instance, some institutions require that a student exceed a particular HSGPA to be eligible to be a Non-Submitter, but don't guarantee admission to those students, whereas others guarantee admission (without regard to test scores) to anyone above a particular GPA. The latter would clearly be considered an Academic Threshold policy, whereas the former could be considered either Academic Threshold or Optional for Some. In addition, some institutions shift from one version of the policy to another—often from the more restrictive "Optional Plus" to the less restrictive "Optional for All"—as they get comfortable with the use and implementation of the policy.

Also note that we included only a small representation of Academic Threshold institutions. As described above, at most Academic Threshold institutions students above a particular HSGPA or class rank are automatically considered without regard to their test scores, but the applicants don't actually make a decision about whether or not to submit their test scores. Therefore, the charts in which we compare "Submitters" to "Non-Submitters" we have typically excluded the Academic Threshold institutions because those students don't actively choose to be Non-Submitters.

For this study, we sought to include institutions of a variety of sizes, levels of selectivity, and geographical locations. We sent initial inquiries to the deans of admission at about 100 TOP institutions, and followed up with those who responded to our initial inquiry. Not all that wanted to participate were able to provide the volume of data we were requesting of them. However, as outlined below, the 28 institutions that are included in our study represent a diverse group of institutions.

Category	Count of Institutions
Control	
Private not-for-profit	24
Public	4
Institution Size Category*	
1,000 - 4,999	17
5,000 - 9,999	7
10,000 - 19,999	1
20,000 and above	3
Carnegie Classification: Basic	
Baccalaureate Colleges	14
Master's Colleges & Universities	7
Doctoral Universities	7
Selectivity	
<30%	4
30%-50%	10
51%-70%	9
>70%	5
Endowment per FTE	
<$20,000	6
$20,001-$50,000	5
$50,001-$100,000	4
$100,001-$200,000	9
>$200,000	4
Geography	
Far West	2
Great Lakes	3
Mid East	8
New England	12
Plains & Southeast	3
URM Enrollment	
<10%	4
10%-20%	15
21%-30%	7
31%-40%	2
US News "National" or "Regional" Ranking	
National	20
Regional	8

Figure 2. Participating Institution Profile. The first seven sections of data drawn from ***IPEDS data 2016***. The last section from USNWR rankings, 2018. *Note, the institutional sizes here reflect total enrollments, including graduate programs, whereas the sizes referred to in our report refer solely to undergraduates.

Among our participating institutions, the proportion of Non-Submitters at each institution ranges widely —from 2% to 52%. Excluding the Academic Threshold institutions, the mean is 21.5% and the median Non-Submitter rate is about 23%.

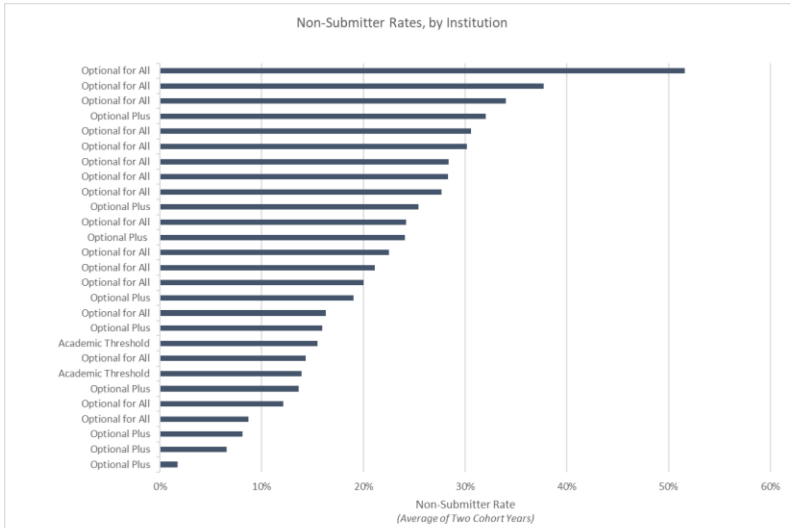


Figure 3. Two-Year Average Applicant Non-Submitter Rate, by Institution.
Exclusion: One institution that did not have Non-Submitter proportions
27[2] institutions **|** 103,088 Non-Submitters | 395,043 Submitters
Enrollment = 1,500-20,000 (2,400 M[3]) **|** Endowment per FTE = $5,000-$800,000 ($65,000 M) **|** Admit Rate = 15%-90% (43% M)

What data was collected and how was it used?

To allow us to assess changes in the admission funnel beginning with the applicant pools, from each institution we received record-level data, consisting of 40 variables (see appendix for detailed list of data definitions) for every applicant from two cohort years prior to adoption of their test-optional policy and two cohort years after adoption of the policy. The exceptions to this form of data gathering were the four institutions that had adopted a TOP prior to 2004 and were unable to submit data from years prior to their adoption of TOPs, and three of the five most recent adopters that could provide data for only one cohort subsequent to their adoption of a TOP.

All data for the study was submitted on a "blind crosswalk" basis, where the institution created a random record ID number for each student record, and only the institution kept that "cross-walk." The study received no identifiable individual information, and we guaranteed confidentiality and anonymity to the institutions in the study, as they shared a great deal of data with us. At the conclusion and publication of this study, each institution can decide if they wish to announce that they were participants.

The following subsections summarize each important area of data collected, and how it was transformed for analysis.

[2] Upon receipt of the data, we learned that one of the "Optional for All" institutions in the study was not able to accurately identify Submitters of testing versus Non-Submitters of testing. Therefore they have been excluded from the analyses in places where those student groups are compared.
[3] The capital "M" stands for Median

A primary focus is on exploring the impact of adoption of a test-optional admission policy on the size and composition of a college's admission funnel – from applicant pool, through admitted students, to enrolled students. Additionally, we have explored the differences between the Non-Submitters and Submitters at each stage of the funnel. This conceptual framework underpinned our approach to gathering, organizing and analyzing the data.

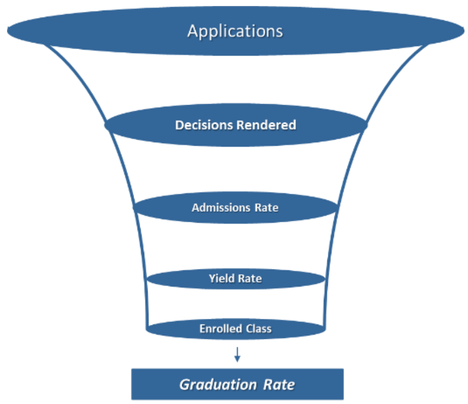


Figure 4. The Admission Funnel Framework. For the purposes of this study the Admission Funnel has been expanded to include a category of **Decisions Rendered** and concludes with a category of **Graduation Rate (**as a measure of student success).

The traditional admission funnel includes prospects, inquirers, applicants, admitted students, and matriculants. For the purposes of this study we collected data beginning with Applicants and added a layer of assessment we have labeled "Decision Rendered" (abbreviated as "DR"), asking the institutions to identify each candidate for whom they actually rendered a decision (admit, deny, waitlist). Thus, the number of DR applicants excludes applicants who remained incomplete and those who withdrew their applications prior to receiving a decision (e.g., made an Early Decision commitment elsewhere). This gave us insight into the inner workings of the policy, for instance, identifying whether Non-Submitters were completing applications at rates equivalent to Submitters.

In addition to the funnel data, institutions provided the following categories of information on each student (details available in the appendix, page 73):

- **Racial and Ethnic Student Identification:** IPEDS-defined categories of racial/ethnic identification.
- **High School GPA Data:** HSGPAs were reported to us as recorded by the institution's Admissions or Institutional Research offices. We did an extensive individual analysis of the GPAs reported by each of the 28 institutions in the study, using whatever scales that institution provided, and allowing for the GPA scales to exceed the traditional 4-point scale. No imputations were used in analysis. For the analyses in this combined report, we truncated all HSGPAs at 4.0.
- **Standardized Test Score Data:** ACT and/or SAT scores were gathered and a concordance table (College Board, 2009) used to convert ACT scores to SAT scores. For simplicity, all references in the report refer to them simply as SATs. The new SAT had not been taken by most[4] of the student cohorts in this study (College Board, 2016). 27% of all Non-Submitter records contained a test score.
- **Financial Aid Data:** The data was categorized into segments (details in the Appendix) using the following financial aid numbers:

[4] Only one institution submitted a few ACT test scores from students post 2016. For this institution, the more recently adopted concordance table was used (though its validity has been disputed by ACT).

- Expected Family Contribution (EFC) – At some institutions this was the federal EFC, and at others it was an institutionally-determined EFC. Because we were attempting to assess aid award differences between Submitters and Non-Submitters at the individual institution, we sought whatever EFC they used to determine the need for aid.
- Total Gift Aid – We requested the total gift aid (from any source) received by the student.
- Pell Grant Recipients – We asked institutions simply to identify any student receiving a Pell Grant, but did not ask for the specific amount of the Pell Grant.
- Merit Aid Recipients – We asked them to identify any student to whom they had awarded non-need-based, "merit" gift aid.

- **Academic Outcome Data:** First-year college GPA (FYGPA), most recent (or final) CumGPA, major designation, a current student enrollment status, and an up-to-date same-school graduation status for all students who enrolled. We used graduation status as our ultimate measure of student academic success. The data was collected in 2016.
- **Test Requiring Institutions, Peer Data:** IPEDS data for both the TOP institutions and their Test Required Policy Peers (TRP Peers) for each of the matched cohorts of students. (For more detail on the selection of Peer institutions, see pages 19 and 76.)

What is the sample size and composition?

We gathered information from 28 institutions, four public and 24 private, of various sizes and levels of selectivity. Of our 28 participating institutions, 20 were new to our research and 8 were repeats from the 2014 study, but with new class cohorts of data. Their undergraduate enrollments ranged from about 1,500 to 20,000, and their admit rates (in their TOP cohort years) range from 15%-90% (15%-80% among the privates, with a median of 43%).

We wanted to examine data from institutions that had varying years of experience with the policy, so there is a range of policy adoption timeframes represented in the study. Four institutions adopted their policies prior to 2004, five adopted their policies very recently, and the remaining institutions adopted the policy between 2004 and 2014. With the exception of the earliest adopters, each of them provided data for two cohorts just prior to adoption of their policies, and two cohorts post policy adoption.

We asked institutions to submit the first cohort ***after*** they adopted the policy, to allow for a full cycle of trial before starting the policy analysis. Seven submitted Post-TOP data beginning the first cohort immediately ***after*** they adopted the policy, and the other institutions submitted data ranging from two to six years after adoption to align data for a 4-5 year graduation rate comparison. Out of necessity, our subset of most recent adopters provided data beginning with the year they adopted the policy.

How was the data analyzed and summarized?

We have a large set of data with many different facets. As noted above, this study was designed not to reach a single conclusion, but to examine and share a wide range of findings. Initially, we did an extensive analysis of the data for each institution individually, searching for common patterns and themes, as well as identifying any apparent inconsistencies that might require clarification by the institution. After these conversations, it was sometimes determined that there was simply missing data or, due to changes in computer systems at the institution, there was inconsistent data across their cohorts. In these instances, we decided to eliminate that particular data element and to exclude the institution from comparisons that were dependent upon that data.

At the institutional level we focused on the differences between Submitters and Non-Submitters at various stages in the funnel. To give an illustrative example of the data gathered, the following chart outlines the funnel activity by the numbers.

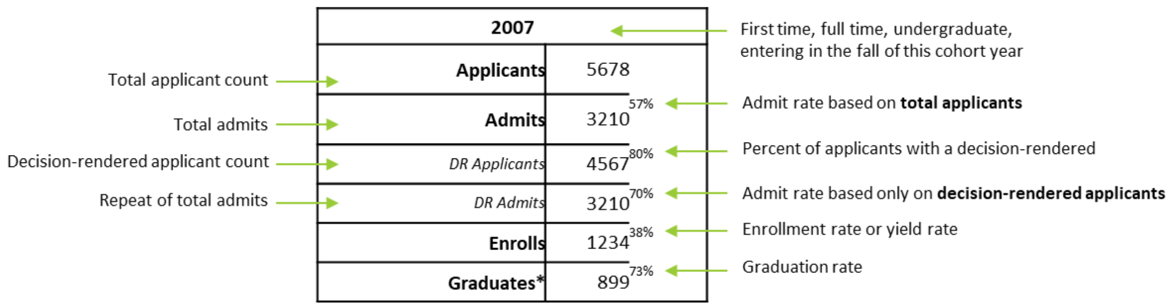


Figure 5. The Admission Funnel – Illustrative Example of Tables. Data requested of the 28 institutions in the study.
**For "Graduates", participating institutions submitted an updated status on all students as of June 2016.*

This table was summarized for each institution for each cohort year submitted, then averaged across pre-policy years, and across post-policy years. It was then filtered by the demographic and admission behavior data we collected, for instance: gender, race/ethnicity, athletic participation, high school type, Early Decision usage, family financial status, and first-generation-to-college status.

In this report, whenever there is a single data point provided for an institution (e.g., admit rate, percentage of URM students, proportion of Non-Submitters, etc.), it represents an average of two cohort years of data (i.e., the two pre-adoption cohorts or the two post-adoption cohorts) whenever that data is available. In some charts, the averages represent averages at each individual institution, whereas in other charts we present the pooled student data using individual records across a subset of institutions (e.g., the graduation rate for Submitters is derived by pooling the number of all the graduates for that set of institutions and dividing it by the total number of students who enrolled in those same cohorts). We have tried to present the data in whichever format provides the clearest meaning for the reader, and in some cases we have provided more than one format.

What are the research limitations?

To summarize, this study is a case study exploration into the role of testing in college admissions. Its very strength is in the attention paid to unique scenarios at each institution, following through to detailed understanding of admission and enrollment patterns. However, these are the limitations that come along with this case study-based approach:

- While we were able to recruit a wide range of institutions to volunteer participation in the study, and we learned quite a bit about each one, the sample size was not large enough to be considered definitively representative of institution or policy type.
- Each institution in the study maintained different record keeping practices, data definitions, and data governance policies. Our very detailed data request form and data transformation approach mitigated many of these differences. However we rarely received identically comprehensive datasets from any two

participating institutions. This led us to declare certain aspects of some data submissions as unavailable, unreliable, or irreconcilable. We have made notes in each Figure description to this effect.

- Working with historical data meant that there were sometimes inconsistencies across cohorts at individual institutions due to changes in computer systems or data coding schema. Similarly, the current staff were often not those who were at the helm during the time of policy adoption. Therefore our understanding of context surrounding the policy (e.g., motivation for adoption, concurrent initiatives, financial aid packaging strategies, and so on) was not always clear.
- The nature of our study inherently dealt with self-selection bias, both at the institution level (e.g., each institution made the decision to adopt TOP) and at the student level (e.g., all but two of the institutions – those with threshold policies -- had policies that allowed students to choose to withhold test scores in the admission process). There are a number of underlying motivations, and demographic and psychographic elements that we cannot account for in this study, rendering the determination of *causation* more difficult.

PRINCIPAL FINDINGS

We open our findings with an homage to the variability of institutions. We present two case studies of institutions that effectively realized the goals they set for their Test-Optional Policy. While trying to provide a clear picture of the institutions, in order to protect their anonymity, we have kept some of the descriptive information broad.

Institution A

Large, urban, private, not-for-profit

This institution described TOP as a means of formalizing what they had been doing all along, to *"Give students a choice about how they might want to present themselves."*

They described the process of transitioning to TOP as largely uneventful, requiring a few more temporary readers to aid the transition because without testing *"you typically need to look more closely at the high school record, the rigor of the curriculum, and the school profile for context."*

After adopting TOP, our data revealed that the new policy drew a smaller than average proportion of students who did not submit testing (9% vs 23%). And, although they increased applications, they grew proportionately more slowly than their matched TRP Peer.

It was a different story for URM students. They enrolled substantially more of these students (***15 percentage points*** more) than they did prior to adopting TOP which was proportionately a 76% greater gain than their TRP Peer. So, in spite of having their applicant pool grow at a slower pace than their TRP Peer, they were able to make substantially greater gains in enrolling URM students (and to a slightly lesser extent with Pell recipients and First-Generation students).

As compared to the pre-policy cohorts, the enrolled TOP cohorts saw a ***12 percentage point rise in the proportion of students with need***, and although this institution did not submit data on gift aid, it likely had to increase aid commitments to support this growth.

Once enrolled, Submitters had a marginally higher CumGPA than Non-Submitters, but both the overall population and the URM population **graduated at virtually identical rates**.

Institution B

Small, suburban, private, not-for-profit

This institution was motivated to adopt TOP because they thought the test-requirement was preventing some students from applying.

After adopting TOP, the applicant pool grew proportionately faster than the TRP Peer (proportionately 40% more growth), with Non-Submitters comprising 19% of the overall pool – a slightly lower proportion than the majority of the institutions in the study.

Their percentage of enrolled URM students *was **17 points higher*** after they adopted a TOP, which was proportionately a 75% greater gain than their TRP Peer. As with most of our institutions, they admitted Non-Submitters at a substantially lower rate than Submitters (15 percentage points lower), and the Non-Submitters enrolled at a substantially higher rate (23 percentage points higher).

At this institution, the post-policy cohorts included a lower proportion of students with need (11 percentage points lower), than the pre-policy cohorts, but ***Non-Submitters were, on average, $4,000 needier***. To the detriment of higher need students, though, this institution seemed to favor low and no-need students in its awarding strategy.

Once enrolled, Submitters had a marginally higher FYGPA and CumGPA (+0.09 and +0.07 respectively) than Non-Submitters. There are mixed results from the two graduating cohorts, with ***Non-Submitters graduating at a higher four-year rate*** (8 points higher) than Submitters, but **lower from the earlier cohort with a five-year graduation rate** (4 points lower). We offer no speculation on the difference other than that perhaps as they refined their review process for Non-Submitters – they got better at it! For both the four-year and five-year cohorts, the ***URM Non-Submitters graduated at a consistently higher rate*** (7 points higher) than the URM Submitters.

DOES A TEST-OPTIONAL POLICY…

Encourage more students to apply?

All institutions saw an application increase, but just over half saw application growth greater than that of a matched test-requiring peer (TRP Peer).

We interviewed the admissions dean at each of the 28 participating institutions about the impetus and the goals behind their institution's adoption of a TOP. They had not all been in their roles at the time the policy was adopted, but the recurring theme was that a major goal had been to increase applications, particularly among underrepresented student groups.

Not surprisingly, our data reveals that participating institutions saw application increases after policy adoption. The overall average increase in applications was 1,926 (and a median of 1,234), with an average increase of 29% at the private institutions in the study and 11% at the public institutions. Note these application increases represent various spans of time from the date an institution adopted a TOP to the date of the latest cohort they submitted for the study. These time spans range from one year to six years, collected during the period of 2004-2016.

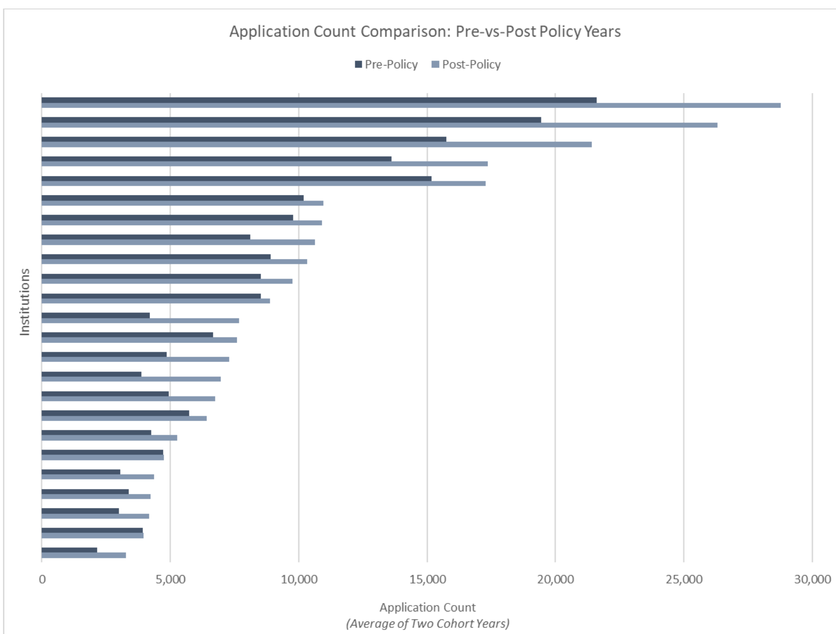


Figure 6. Pre-Policy vs. Post-Policy Comparison of Average *Application* Counts, by Institution.
Exclusions: Four policy early adopters
24 Institutions | 888,021 Records
Enrollment = 1,500-20,000 (3,000 M[5]) **|** Endowment per FTE = \$4,000-\$500,000 (\$63,000 M) **|** Admit Rate = 20%-80% (50% M)

These universal increases are not surprising, as most institutions posted gains during this period. The critical question is whether our TOP institutions fared better or worse than peer institutions that still required standardized tests of

[5] Going forward, "M" represents the Median

their applicants. A 2015 study out of the University of Georgia (Belasco, 2014) sought to answer this question and concluded that, in the aggregate, selective liberal arts colleges that adopted a TOP (in comparison with those that continued to require SAT or ACT scores) had not increased their enrollment of URM students or Pell recipients. But in focusing on the averaged outcomes, that study may not have been able to discern impact at the institutional level.

Recognizing that there is great variation among institutions in selectivity, size, financial resources and geographical markets, and that not all test-optional admission policies are the same, nor pursued with the same vigor, it seemed appropriate to apply a more institutionally-focused approach to answering this question. As noted earlier, we asked the admissions dean from each of our participating institutions to identify their top three "peer competitors" – the institutions they felt were most like their institution, and that were in most direct competition for their students (i.e., not an "aspirational" institution, but one with which they had a fairly even win-loss ratio with students admitted to both institutions). Then we identified institutional match criteria to finalize the selection of the single, most comparable, Test-Required Policy (TRP) Peer match[6] for use in our analysis.

Comparisons are made using data from the same cohort years for the Test-Required Peer as were submitted to us by the Test-Optional institution, but for these comparisons, all data for both institutions was drawn from IPEDS to ensure consistency in the comparison.

The following chart illustrates whether each TOP institution experienced a greater or lesser percentage gain in applications than their Test-Required Peer institution. We analyzed the application growth by using percentage growth (rather than raw numbers) to compensate for the varying sizes of institutions (enrollments range from approximately 1,500 to about 20,000[7]). Thus, an applicant pool of 1,000 that increases to 1,100 would be represented as having growth of 10%, and so would an applicant pool of 10,000 that increases to 11,000). The differences represented below are the proportionate differences between the percentage growth of each TOP institution and its Test-Required Peer. If a TOP institution experienced growth of 22% and its TR Peer experienced growth of 20%, the proportionate difference represented below would be 10% (22%/20%) greater proportionate growth for the TOP institution.

In this chart we observe that 13 of 23 (57%) of the TOP institutions in our study experienced greater proportionate growth in overall applications than their TRP Peers during the same time period, while only six (26%) of them experienced less application growth than their TRP Peers. Four experienced essentially the same level of growth (within +/- 3%).

[6] Data was pulled from IPEDS Data Center. The following criteria was used to select (from the list of peer-competitors identified by the dean) the best possible Test-Required Peer: 1) Control type - Public or Private, 2) Similar URM proportion during the TOP pre-policy cohort years, 3) Similar Pell proportion, or if unavailable, similar federal aid award proportion, during the TOP pre-policy cohort year, 4) Similar application pool size, 5) Same general geography, 6) Carnegie Classification: Size and Setting.

[7] IPEDS enrollment data on undergraduates entering Fall 2016.

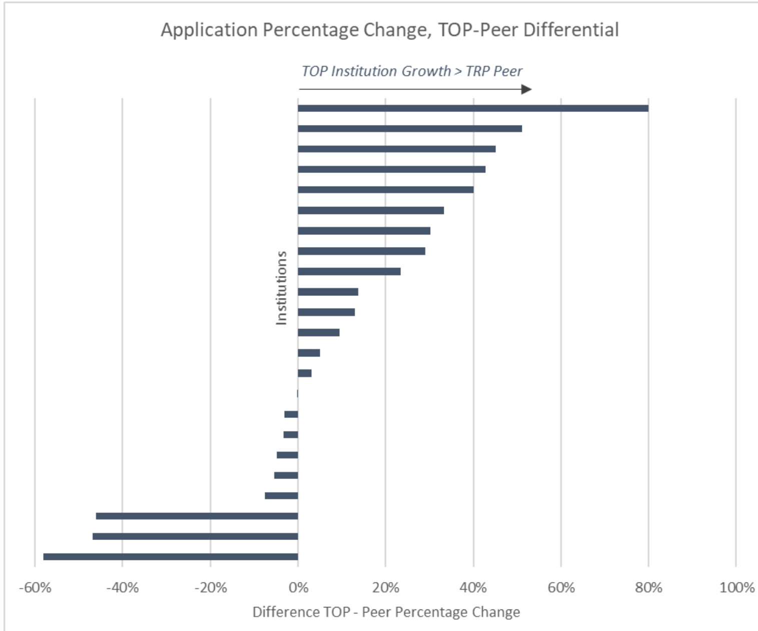


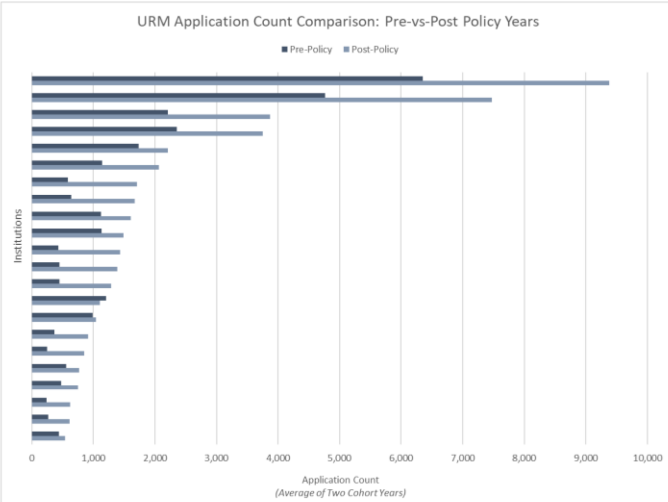
Figure 7. Application Change Differential, TOP Institution versus TRP Peer During Pre-and Post-Policy Cohort Years. IPEDS data on corresponding, averaged, pre-policy and post-policy cohort year data on FTFT degree-seeking undergraduates.
Exclusions: Four TOP institutions and respective matches were excluded due to lack of data available prior to TOP adoption (i.e., policy early adopters), and one additional TOP was eliminated due to lack of a well-matched TRP Peer.
23 TOP participants and matching TRP peers (N=46) | 1,164,546 Applicant Records

> *"The policy has worked, though it is not nearly as popular (widely used) as we thought it would be…maybe most of the students who would traditionally have been attracted to TOP were already applying without being overly concerned if their test scores didn't represent them well"*
>
> *Dean from large private university*

Does a TOP help institutions enroll more traditionally URM and low-SES students?

For the majority of our TOP institutions, yes. Almost all institutions saw a rise in URM applications after TOP, and two-thirds of them saw that rise correspond to URM enrollment growth above that seen by a matched TRP peer institution. Half experienced Pell enrollment growth above a TRP peer.

Based solely on the raw counts of URM applicants and enrollees, the data shows that ***all but one*** of our TOP institutions experienced an increase in URM applications after TOP adoption. The overall average increase in application count was 835, with a slightly higher median of 938. Similarly, all but three of them increased their enrollment of URM students after adoption of a TOP.



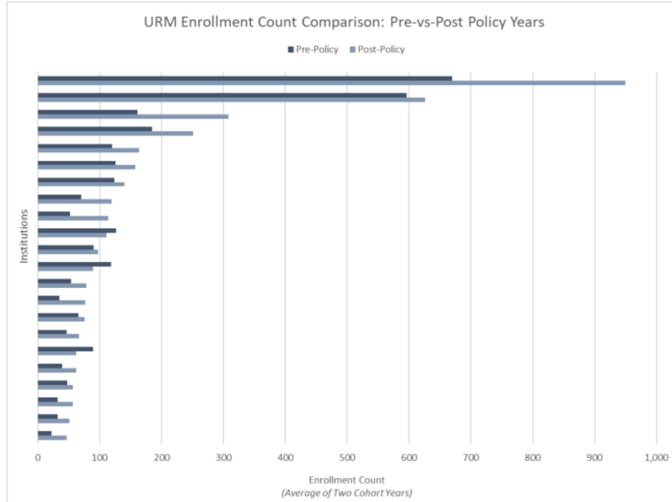
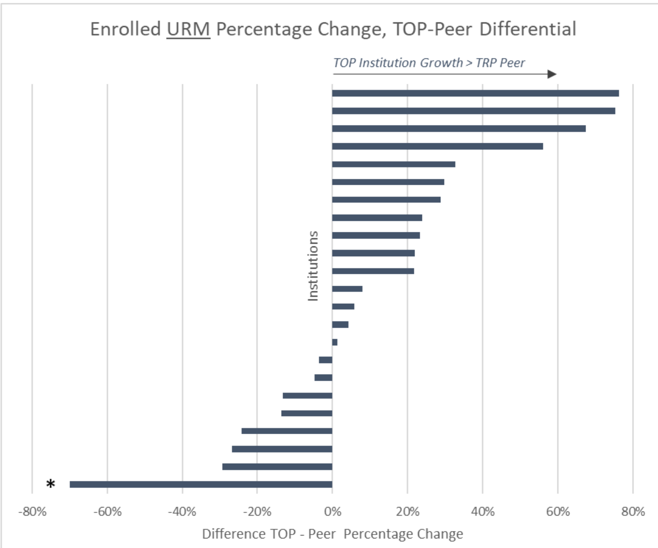


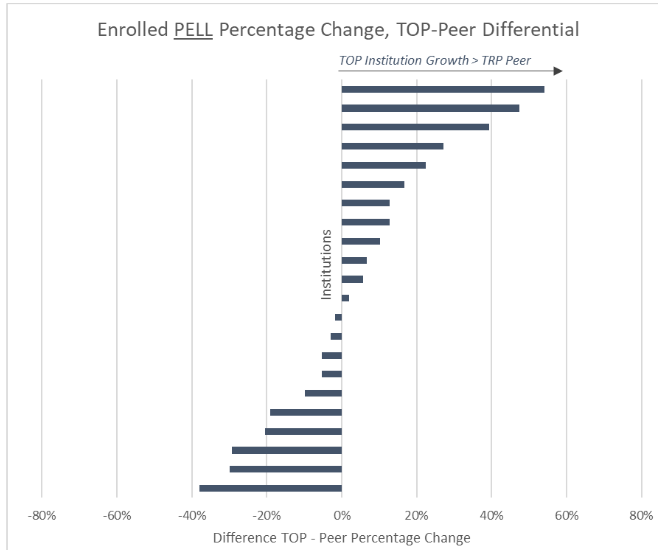
Figure 8. Pre-Policy vs. Post-Policy Comparison of Average URM *Application* and *Enrollment* Counts, by Institution.
Exclusions: Four policy early adopters and two institutions with unreliable URM data
22 Institutions **|** 74,770 URM Applicants **|** 13,613 URM Enrolled
Enrollment = 1,500-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$500,000 ($60,000 M) **|** Admit Rate = 20%-80% (46% M)

> *"I looked at the average test scores of colleges, and if my scores didn't fit that range, I just nixed them. That was the first time that I was presented with the idea that SATs could be optional; I didn't know. I would never have thought to apply to liberal arts schools in the Northeast. "*
>
> *Hispanic female from rural Texas*
> *in "The Test and the Art of Thinking" (Davis, 2018)*

It is important to try to assess whether offering the option to apply as a Non-Submitter actually increased the proportion of underrepresented students who chose to apply and enroll at an institution, or did it simply shuffle the deck, having no real impact on the composition of the class? Given that this timeframe coincided with a period of rapid growth in the diversity of college applicants, we used the matched TRP Peers to assess any true differences. The differences in proportions in the enrolling classes are represented below in a manner parallel to Figure 9.

There were 14 of 23 (61%) of the TOP institutions that achieved proportionately greater increases in enrollment of URM students than their TRP Peers; one was essentially the same (within +/- 3%); and eight (35%) fared less well than their TRP Peers. Enrollment of Pell Recipients was more evenly split, with 11 of 22 (50%) of the TOP institutions increasing the proportion of Pell recipients more than their TRP Peers, three enrolling roughly the same proportions and eight (36%) losing ground in comparison with their TRP Peers.





**This matched pair was the only one in which the URM proportions for all reported classes for both the TOP and TRP Peer were 7% or less, so this representation should be viewed with caution as the numbers are small.*

Figure 9. Enrolled URM and Pell Proportion Change Differentials, TOP Institution versus TRP Peer During Pre-and Post-Policy Cohort Years. IPEDS data on corresponding, averaged, pre-policy and post-policy cohort year data on FTFT degree-seeking undergraduates. Institutions are <u>not</u> aligned across the two charts.
23 TOP participants and matching TRP peers (N=46)[8]

Finally, to summarize the pre-to-post policy analysis, we completed a statistical test to compare the institutional averages between our TOP institutions and their TRP Peers. This statistical testing, as outlined below, indicates that our TOP institutions experienced greater application and URM enrollment growth than their matched peer institutions. As seen below, using a statistical measure of effect size (Cohen's d) we find a ***medium effect size*** between the proportionate differences in the mean application growth and the mean URM enrollment growth for the TOP institutions vs their matched TRP Peers. There is only a small effect size seen for Pell recipient enrollment proportions. (Because the data was drawn from IPEDS, we could not compare growth in the enrollment of First Gen students.)

[8] Note: In the Pell comparison, one additional TOP and its match were excluded because policy adoption occurred too recently for IPEDS financial aid data to be available. Prior to 2007 when Pell proportions were not available in IPEDS, proportion of "students awarded federal grant aid" was used.

TOP INSTITUTION VS TRP PEER – COHEN'S D STATISTICAL COMPARISON OF MEANS				
		N	Mean	Cohen's d
APPLICATION % CHANGE	TOP Institution	23	.285	Medium Difference (.4)
	TRP Peer	23	.179	
URM PROPORTION ENROLL % CHANGE	TOP Institution	23	.344	Medium Difference (.4)
	TRP Peer	23	.217	
PELL PROPORTION ENROLL % CHANGE	TOP Institution	22	.253	Small Difference (.1)
	TRP Peer	22	.210	

Color Key	
No Difference	< 0.1
Small Difference	0.1 - 0.3
Medium Difference	0.3 - 0.5
Large Difference	> 0.5

Figure 10. Effect Size TOP vs TRP Policy Comparison for **Overall Applicant**, **Enrolled URM**, and **Enrolled Pell**. TOP Institution versus TRP Peer During Pre-and Post-Policy Cohort Years. IPEDS data on corresponding, two-year average, pre-policy and post-policy cohort year data on FTFT degree-seeking undergraduates.
23 TOP participants and matching TRP peers (N=46)

As a reminder – the charts above are measuring the difference (between the TOP institution and its TRP Peer) in the amount of "improvement" on each characteristic. But as indicated above, all the TOP institutions had increases in the actual number of applications. All but one received an increased number of applications from URM students. All but two enrolled more URM students. And all but one enrolled more Pell recipients. So, the institutions at the bottom of each chart didn't fall behind, they just didn't advance as much as their TRP Peer institution.

Some additional observations in the comparisons above caught our interest:

- The institution that had the least growth (in comparison to its TRP Peer) in applications, also had the least growth in Pell recipients, but had among the strongest growth in URM enrollment, suggesting that this institution many have focused its recruitment and enrollment efforts on this population.
- In a similar flip-flop, the institution with the lowest comparative growth in URM enrollment had among the highest comparative growth in applications overall.
- Three of the four public institutions in the study were included above, and it is interesting to note that all three of them were among the eight that increased *less than* their TR Peers in enrolling URM students.

Worth noting, a small subset of institutions repeatedly appeared as outliers "in the negative" compared to their TRP peer in the above charts. A later section of the report (p. 46) will provide some additional perspective on these cases.

So it is clear that, in comparison to their TRP Peer institutions, TOP institutions have varied outcomes relative to the characteristics we assessed. However, the overall comparison suggests a positive relationship between TOP policy adoption and application growth, URM enrollment growth, and slightly less so, Pell enrollment growth.

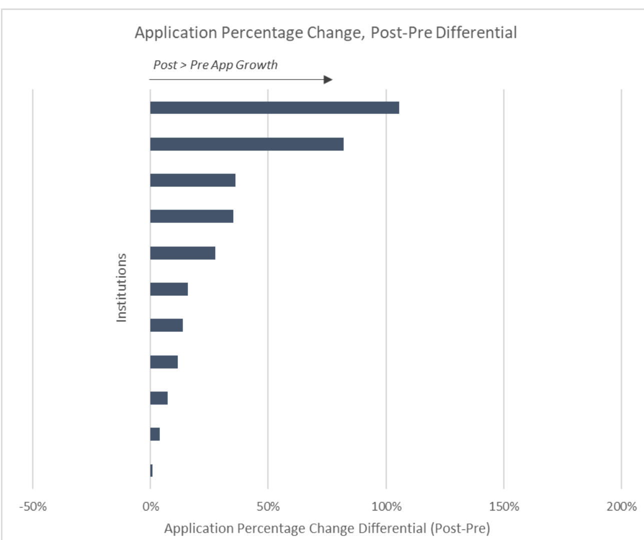
Does a TOP negatively impact the patterns seen in admissions, from selectivity to enrollment rates?

The answer is no. All of the institutions that provided consecutive cohort years pre-policy to post-policy data saw overall application growth, and all but one saw URM application growth. A few institutions saw an increase in selectivity.

A major concern about making a significant change in an admission policy is whether doing so will in some way have a negative impact on the quantity, quality, or composition of the applicant pool and, ultimately, on the enrolled student body.

The final four charts in this section include information from the 13 institutions that submitted data from the cohorts immediately preceding and immediately following their adoption of a TOP. While we cannot isolate the impact of the adoption of the policy from the impact of other changes occurring concurrently, by limiting this comparison to these institutions, we were able to observe the changes that were synchronous with the policy adoption.

At the applicant stage we see that all of these TOP institutions had increases in the number of applications ranging from trivial to a doubling of apps in the five-year period. And all but one of them experienced substantive gains in the number of applications submitted by URM students.



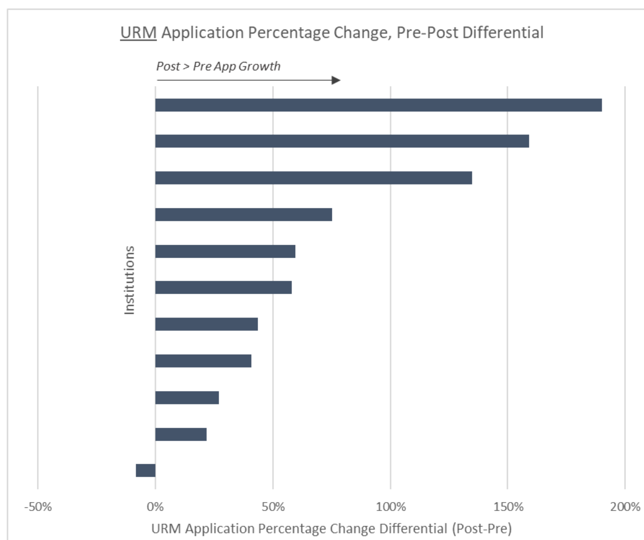


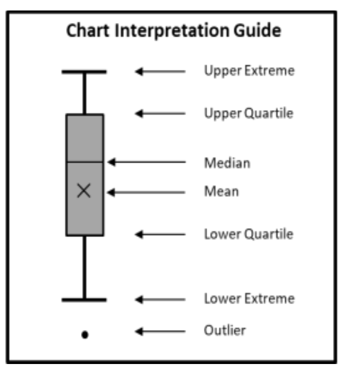
Figure 11. Pre-Policy to Post-Policy Growth of URM Applications, by Institution. The institutions represented include thirteen public and private TOP institutions in the study that provided cohorts *immediately preceding and immediately following their adoption of a TOP*.
Exclusions: Four policy early adopters and two institutions with unreliable URM data
All Applicants: 13 Institutions **|** 656,491 Records
URM Applicants: 13 Institutions**|** 138,482 URM Records
Enrollment = 2,000-10,000 (4,000 M) **|** Endowment per FTE = $4,000-$250,000 ($60,000 M) **|** Admit Rate = 20%-80% (45% M)

"After several years of essentially no growth in African-American enrollments, our first year of TOP had a dramatic increase in African-American, Hispanic, and International apps."

Dean from large private university

Although it is impossible to know what would have happened if these colleges had <u>not</u> adopted a Test-Optional admission policy, the charts below suggest that their applicant pools have not suffered subsequent to adopting the policy. The first chart compares aspects of the funnel for all students in the cohorts during the pre-policy (Test-Required) years against those of the ensuing post-policy (Test-Optional) cohorts.

Thus, it appears that for this group of colleges, the decision to adopt a TOP has not had a negative impact on their admission funnels. To illustrate the impact, we chose a "box and whiskers" style of chart for a number of comparisons as it provides a multi-dimensional visual representation that allows the reader to simultaneously view the complete range, the middle 50%, the mean and the mode, as well as any outliers, thereby illustrating the sometimes-wide variation between institutions and their experiences with TOP. The following outlines a brief guide to interpretation of the subsequent charts:



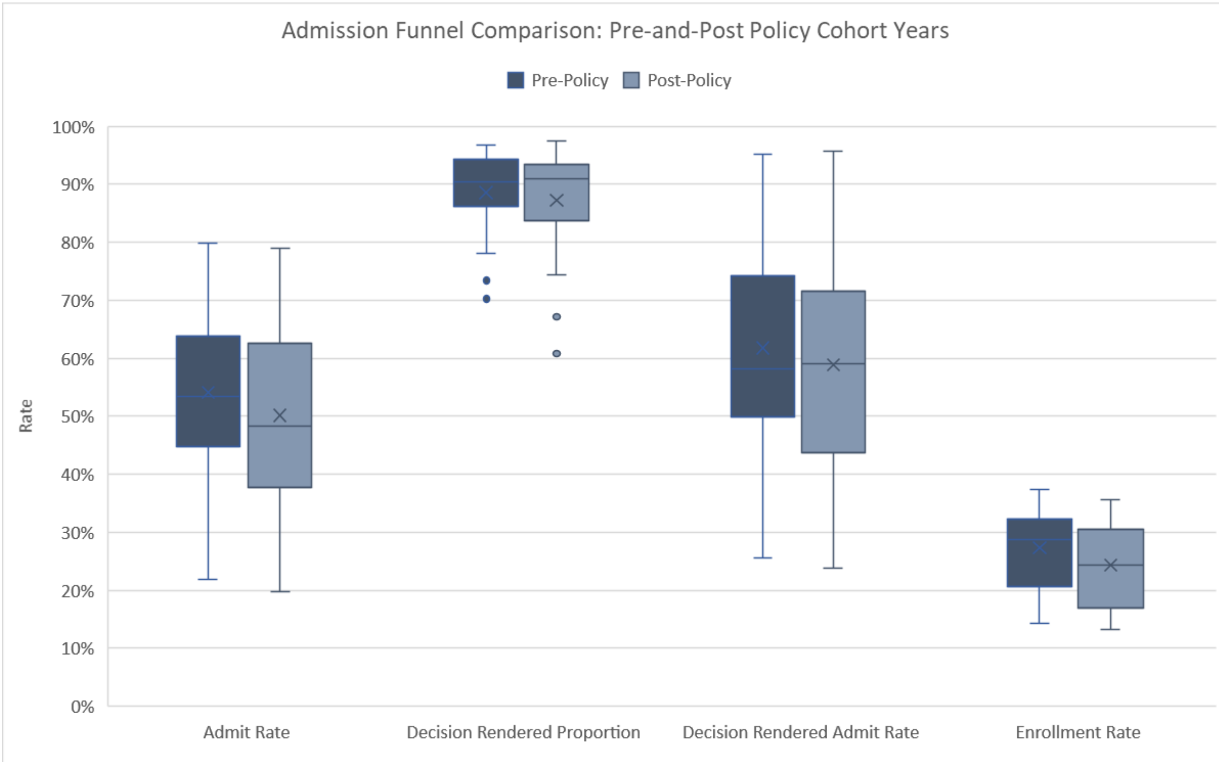


Figure 12. Pre-Post Policy Funnel Overview. Cohort Years 2008-2016. The institutions represented include thirteen public and private TOP institutions in the study that provided cohorts immediately preceding and immediately following their adoption of a TOP.
Exclusions: Four policy early adopters
13 Institutions **|** 656,491 Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$250,000 ($60,000 M) **|** Admit Rate = 20%-80% (50% M)

The funnel patterns of the Pre-Policy cohort years are very similar to the Post-Policy cohort years. The mean and median Admit Rates are marginally lower, as are the Enrollment Rates (yield). But as applicant pools increase in size, it is not unusual for the institutions to become somewhat more selective. Similarly, yield rates at colleges have tended to decline over the past couple of decades as students, on average, have applied to an increasing number of colleges.

One study offered the opinion that colleges were becoming test-optional, not to increase diversity, but to appear more selective (Belasco, 2014). As readers will see later, the admit rate for Non-Submitters is modestly lower, but in the chart above the modest overall differences in admit rates from pre-to-post TOP do not offer much credence to the argument that colleges are recruiting Non-Submitters only to turn them down.

The next view of the funnel—focusing exclusively on URM students—displays similar, but more exaggerated patterns. The median Admit Rate and the quartile span during the test-optional years is lower and wider —from 16% to almost 60% -- compared to the test-required years. The Decision Rendered proportion is equivalent, but with some significantly lower institution outliers. However, the enrollment rate quartile range has a distinctly wider span, indicating that some institutions saw a significant rise in yield, while others experienced a drop.

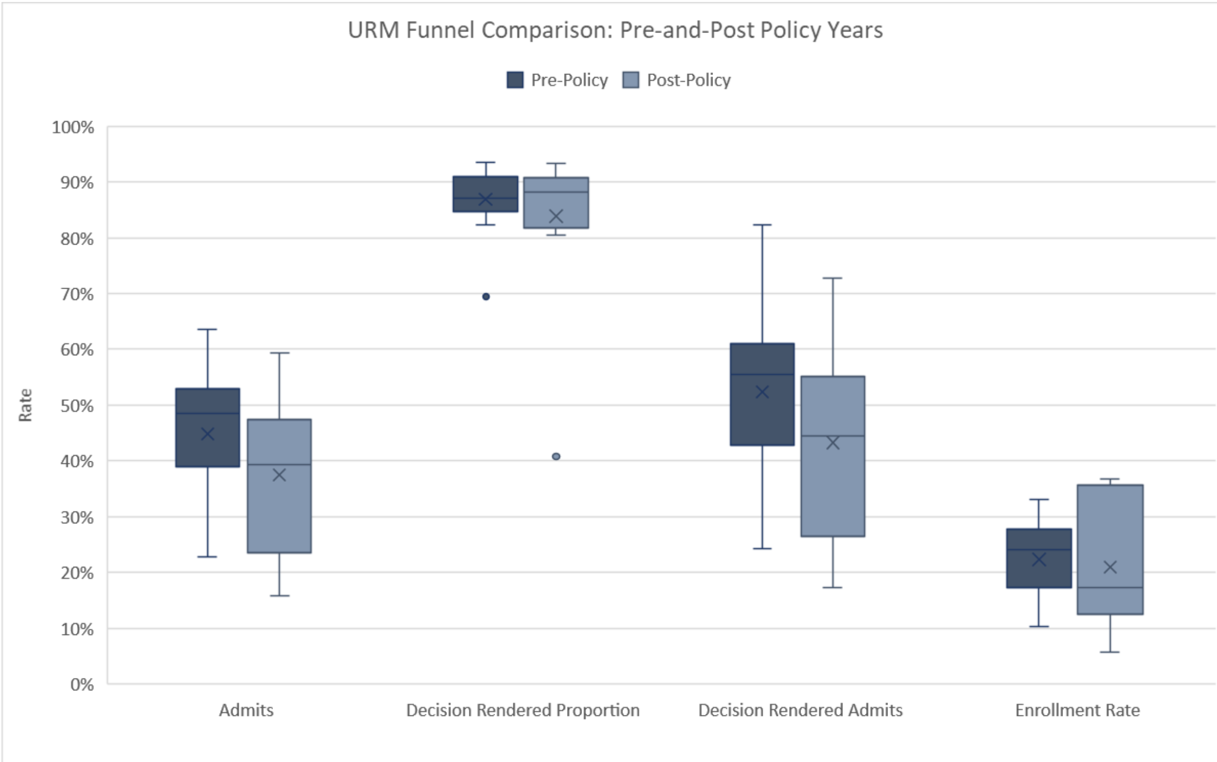


Figure 13. Pre-Post URM Admission Funnel. Cohort Years 2008-2016. The institutions represented include thirteen public and private TOP institutions in the study that provided cohorts immediately preceding and immediately following their adoption of a TOP.
Exclusions: Four policy early adopters, and two that did not have reliable URM data
13 Institutions **|** 138,482 URM Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = \$4,000-\$250,000 (\$60,000 M) **|** Admit Rate = 20%-80% (45% M)

In summary, as noted elsewhere, there is great variation among the experiences of colleges that have adopted test-optional admission policies. And while it is seductive to believe one can make a single pronouncement about the impact of adopting a TOP – much of that impact varies based upon the specific institution, its competitive position in the world of higher education, and the implementation and promotion of the test-optional policy. Our participating TOP institutions varied in size (~1,500 to ~20,000) and selectivity (with admit rates ranging from ~15% to ~90%). But the experiences of this particular batch of colleges suggests that the adoption of a well-promoted and well-executed test-optional admission policy can reasonably lead to an increase in overall applications as well as an increase in the URM representation (both numeric and proportionate) within the freshman class. As such, a TOP policy can provide one tool to assist a college in attracting and enrolling a larger contingent of URM students.

NON-SUBMITTER PROFILE

Do Non-Submitters and Submitters exhibit different funnel patterns?

The answer is yes. Non-Submitters are admitted at lower rates, but enroll at significantly higher rates than Submitters. Non-Submitters go on to graduate at rates equivalent to Submitters.

In addition to assessing the broad impact of a Test-Optional admission policy on an institution's applicant pool and enrolled classes, this study sought to identify any differences between the funnel patterns of the students who submitted test results (Submitters) and those who chose not to submit test results (Non-Submitters) in the admission process. The chart below illustrates the differences between these two groups at the various stages of the admission funnel.

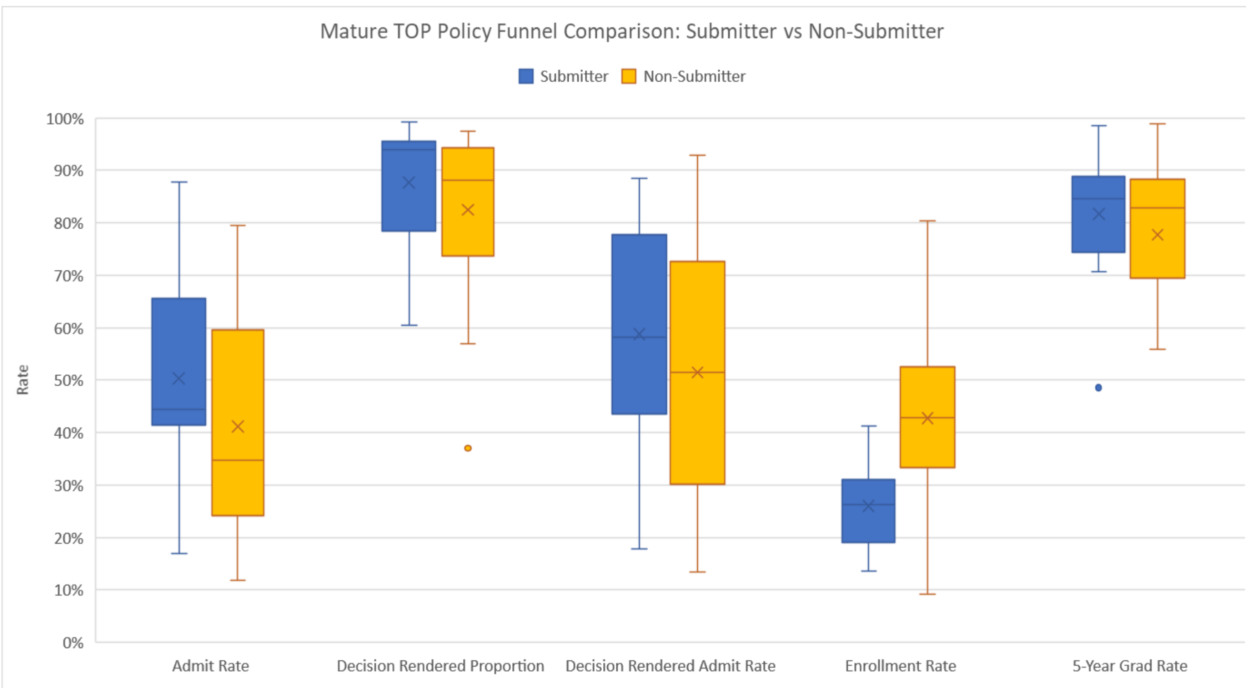


Figure 14. Mature TOP Policy Admission Funnel, Submitter vs. Non-Submitter Comparison, with 5-Year Graduation Rates. Data represents 14 public and private institutions in the study for which we had 5+year graduation data.
Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions
14 Institutions | 166,561 Records
Enrollment = 1,500-5,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-80% (45% M)

There are several notable differences, but perhaps the most significant are that, on average the 14 institutions in the chart above (those that had at least 5-yr graduation rates available) admitted Non-Submitters at lower rates than Submitters, and, on average, those Non-Submitters enrolled (yielded) at substantially higher rates, and went on to graduate at similar rates. The graduation rates from Non-Submitters at these mature TOP institutions have a wider range and lower mean than seen among Submitters, however the differences between the two groups' averages and medians are within 3-4 percentage points. It is important to note that this data represents *institution averages*. As we will illustrate later, pooled student data on average graduation rates shows a comparable, but slightly different picture.

Using the same lens, the chart below focuses on the admission funnel for underrepresented minority (URM) students (N.B. the chart below excludes two institutions that were not able to present reliable URM data at the admit stage.) The URM funnel mimics the patterns seen in the All Student funnel -- institutions admitted URM Non-Submitters at lower rates than URM Submitters, but the URM Non-Submitters enrolled at significantly higher rates— an average of 14 percentage points higher than Submitters. URM graduation rates are harder to reliably interpret, because graduated Non-Submitter URM counts are low, but the URM 5-year graduation rates show equivalence between the two groups.

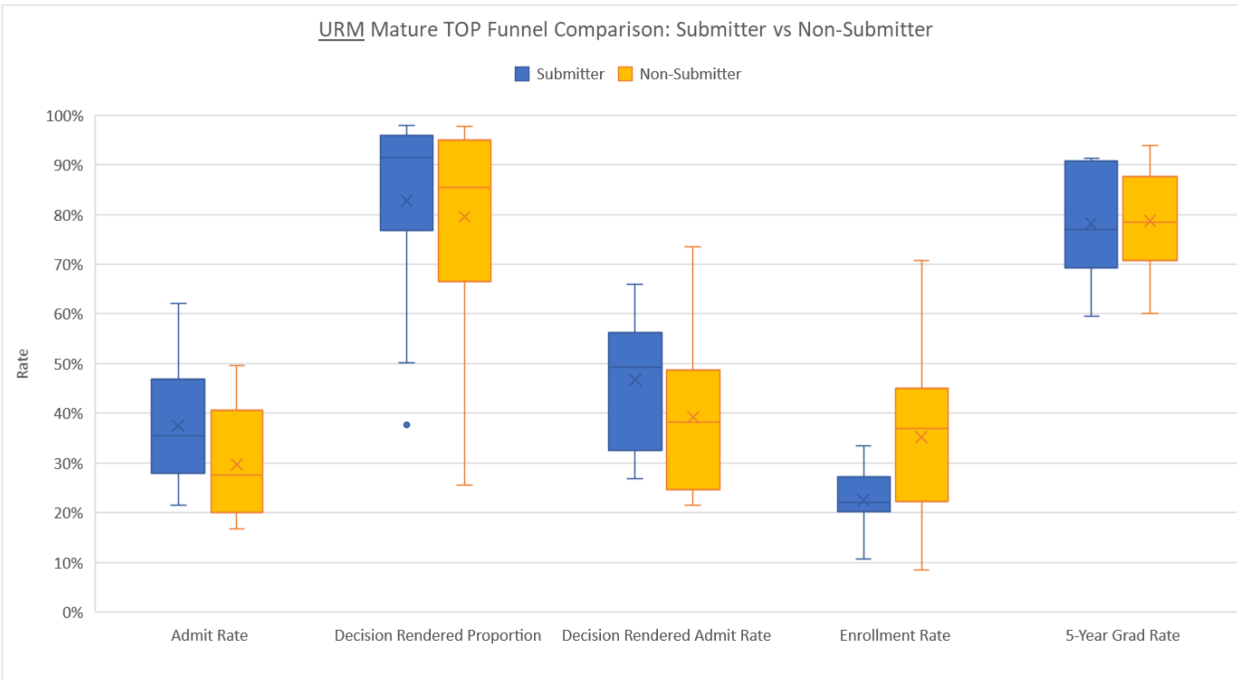


Figure 15. Mature TOP Policy *URM* Funnel, Submitter vs. Non-Submitter Comparison, with 5-Year Graduation Rates. Data represents 12 private TOP institutions with reliable URM data for Submitters and Non-Submitters and 5+Year graduation rates.
Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions, and two that did not have reliable URM data. No public institutions had reliable data for this assessment.
12 Institutions | 26,245 URM Applicant Records
Enrollment = 1,500-5,000 (2,500 M) **|** Endowment per FTE = $44,000-$800,000 ($150,000 M) **|** Admit Rate = 15%-70% (40% M)

The chart below includes the 8 institutions that adopted TOPs more recently, and thus do not yet have graduation rates to report. But the patterns are similar – they admitted Non-Submitters at markedly lower rates, and Non-Submitters enrolled at higher rates.

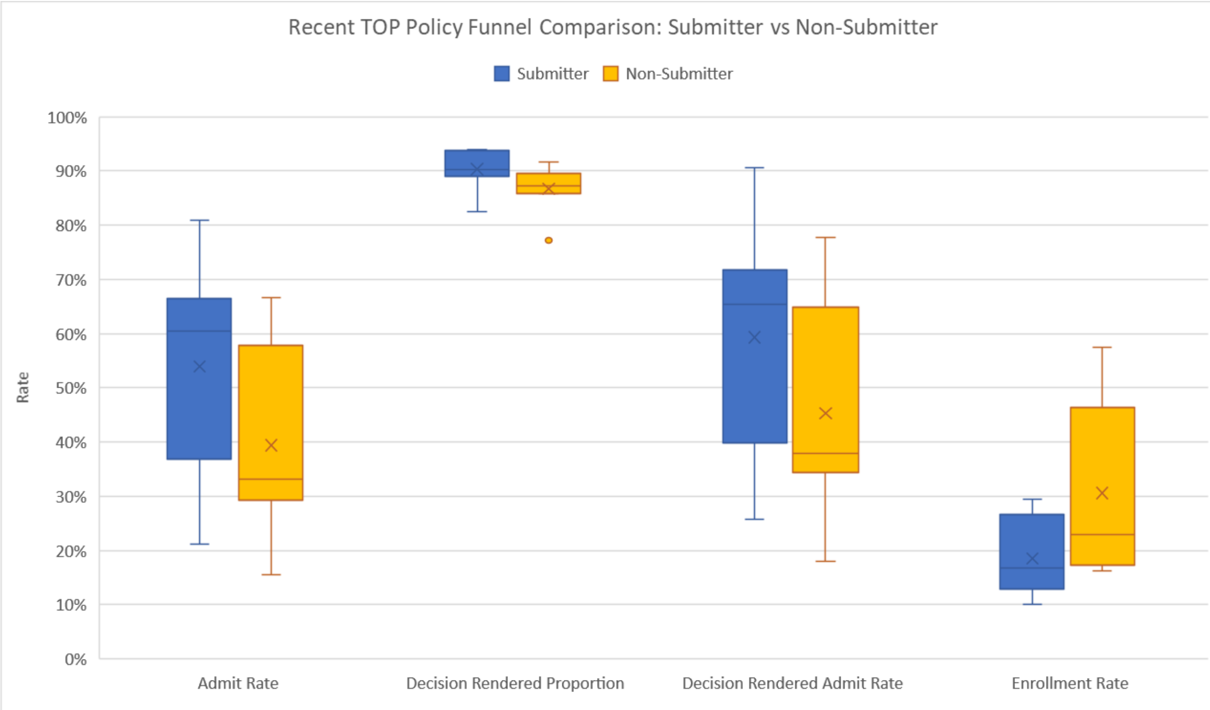


Figure 16. Recent TOP Policy Admission Funnel, Submitter vs. Non-Submitter Comparison, with 5-Year Graduation Rates. Data for chart represents 8 public and private institutions that adopted a policy between 2013 and 2016, and therefore do not have graduation rates to report. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
8 Institutions | 246,565 Applicant Records
Enrollment = 2,000-20,000 (3,000 M) **|** Endowment per FTE = \$15,000-\$250,000 (\$70,000 M) **|** Admit Rate = 20%-80% (50% M)

Figure 17 below, an institution-by-institution comparison of differentials, shows that the majority of these institutions (21 out of 25) admitted Non-Submitters at lower rates than Submitters. However, all but one of these institutions (24 of 25) saw Non-Submitters enrolling at higher rates than Submitters.

Note that shaded bars identify the public institutions.

Figure 17. Institution Admit and Enrollment Rate Differentials: Non-Submitter vs. Submitter. 25 public and private institutions with reliable Non-Submitter data. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
25 Institutions | 396,921 Decision Rendered Records
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($70,000 M) **|** Admit Rate = 15%-80% (40% M)

In a few instances, colleges identified as "Optional Plus" that placed additional requirements on Non-Submitters (e.g., a required interview or written responses to additional questions) may have increased their gross number of Non-Submitter applicants, but had a lower average completion rate, thereby depressing the number of their Decision Rendered (DR) Non-Submitters. In a couple of cases, after adopting a TOP, an increase in URM apps did not carry through to an increase in the number admitted, because there was a large tail-off in the number of URM that received a decision (which we assume was due to lack of completed apps). As illustrated below, a lower proportion of the applicants to "Optional Plus" institutions than to "Optional for All" institutions (78% vs 90%) actually received a decision.

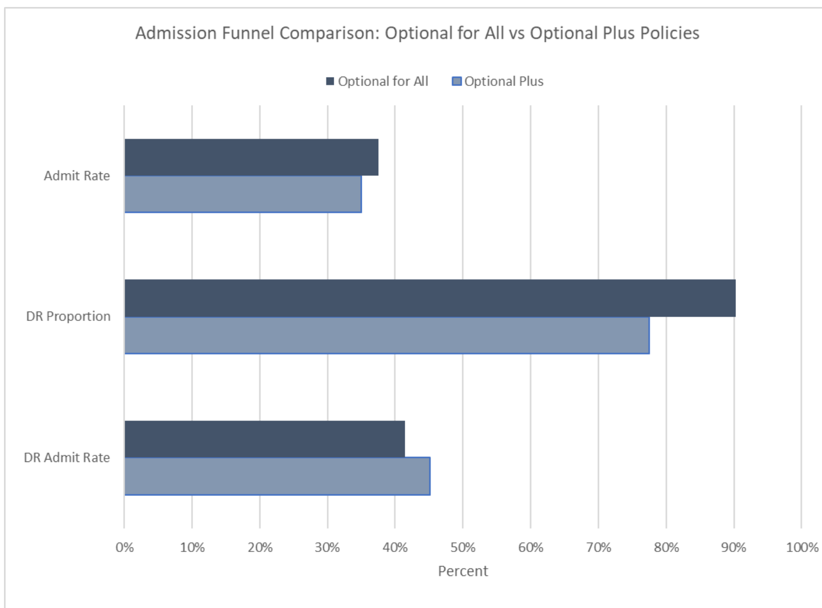


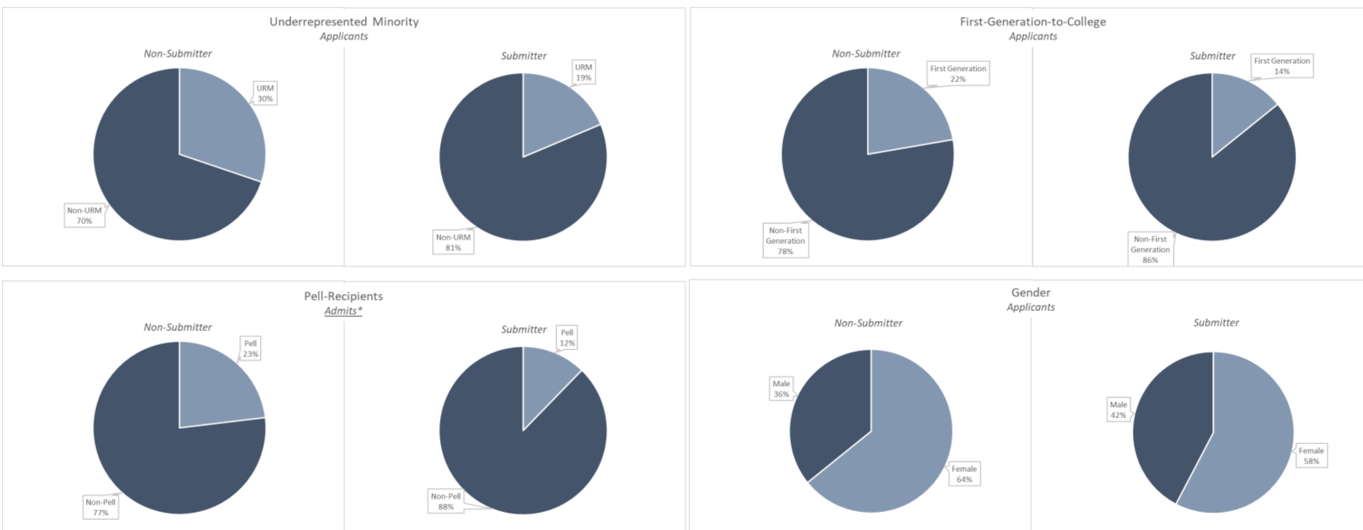
Figure 18. Policy Comparison, by Phase of the Funnel. **Pooled Student Data.** The chart represents data from the 25 public and private TOP institutions with the appropriate policies. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions
25 Institutions **|** 479,008 Records
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($70,000 M) **|** Admit Rate = 15%-80% (40% M)

Who are Non-Submitters of testing?

As in the 2014 "Defining Promise" study, underrepresented minorities, First-Generation students, women and Pell Recipients are more strongly represented among Non-Submitters*. Black/African-Americans chose to be Non-Submitters at higher rates than other racial/ethnic groups.*

When given the opportunity, who chooses to be a Non-Submitter? Based on just under 500,000 records of students applying to these 24 colleges under a test-optional policy, we found that some important subgroups of students stood out as using the policy at higher rates than other student subgroups.

As found in our prior study, "Defining Promise," the Non-Submitter student group included larger proportions of URM students, First-Generation students, and Pell Recipients than seen in the Submitter group. Similarly, women chose to be Non-Submitters at higher rates than men.



**Please note that Pell data is not available at the Applicant stage, so this proportion represents Admits.*

Figure 19. Percentage of Select Student Demographic Segments, Non-Submitter vs. Submitter Comparison. **Pooled Student Data.** Each set of charts represent data from a subset of institutions that provided reliable data. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
URM: 24 Institutions **|** 470,273 Records (of which 99,298 are URM)
First-Generation-to-College: 22 Institutions **|** 384,703 Records (of which 62,626 are First Gen)
Pell: 16 Institutions **|** 110,901 Records (of which 16,016 are Pell Admits)
Gender: 23 Institutions[9] **|** 379,605 Records (of which 224,975 are Female)
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $6,500-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (50% M)

---

[9] Note that two of our larger institutions did not submit gender data, bringing our count down to 23. An "other" gender category was offered, but there was very limited data to represent, with the exception of one institution where virtually all students were recorded as "other." That institution has also been excluded from this average.

The pie charts above compare the proportions of the Non-Submitters and Submitters that were members of each of the designated subgroups. The table below displays the distribution of the applicants based on racial and ethnic status. Although Whites have the largest representation among both Submitters and Non-Submitters, we can see here that Hispanic and Black students both had substantially higher proportionate representation among the Non-Submitters than among the Submitters.

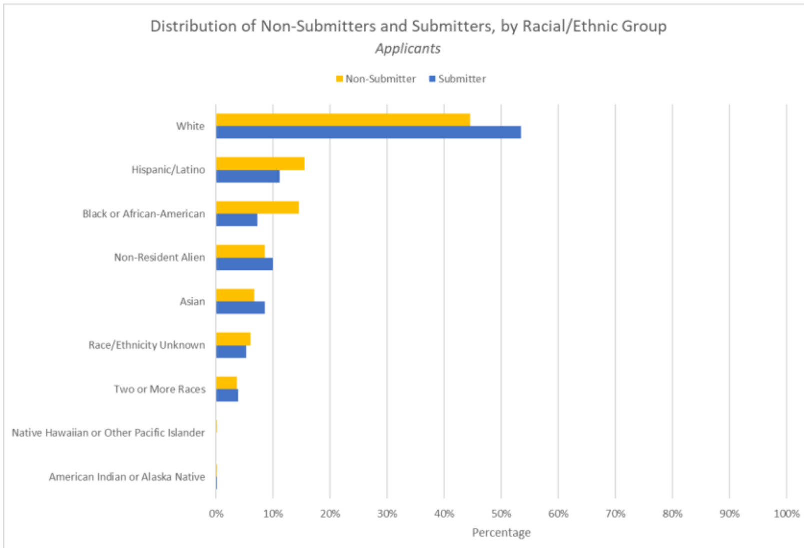


Figure 20. Distribution of Non-Submitters and Submitters by IPEDS Racial/Ethnic Student Group. **Pooled Student Data.** Twenty-four institutions provided reliable Submitter and Non-Submitter URM data.
Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions, and two with unreliable URM data.
24 Institutions **|** 470,273 Records
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $6,500-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (50% M)

As illustrated in Fig 20 above, the cadre of students choosing to apply as Non-Submitters has disproportionate representation from the Black and Hispanic groups. However, the reader should recall that it is not the case that Non-Submitters comprise a majority of any of these groups. So, for instance, while a larger proportion of the Non-Submitters are Black students, among all the Black students, 38% applied as Non-Submitters, as seen below in Fig 21.

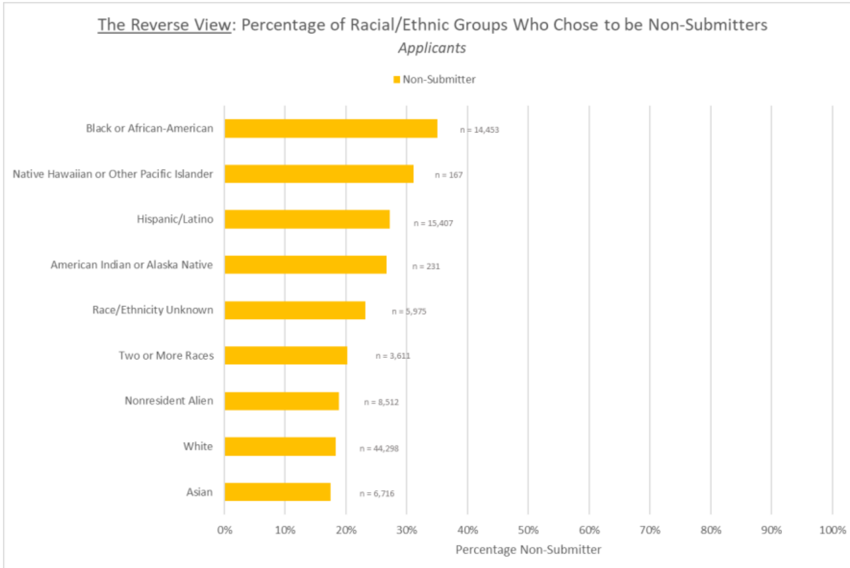


Figure 21. Percentage of Applicants Who Chose to be Non-Submitters by IPEDS Racial/Ethnic Student Groups. **Pooled Student Data.** Twenty-four institutions represented with reliable Submitter and Non-Submitter Racial/Ethnic applicant data.
Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions, and two with unreliable URM data.
24 Institutions **|** 99,370 Non-Submitter Records
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $6,500-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (50% M)

In the figure above, we were somewhat surprised to find almost 20% of non-resident aliens listed as Non-Submitters. In the early years of TOP, almost all non-resident aliens were required to submit testing, on the (perhaps flawed) assumption that institutions needed these tests as evidence of English fluency. It does appear that some international students are now being permitted to apply as Non-Submitters, perhaps from schools with English as the language of instruction, or because American international admissions officers know schools abroad far better, or perhaps some of these are undocumented students. But we are not completely confident of this Non-Submitter statistic for internationals. It may be that some students are submitting TOEFLS, IBs or other tests which are not being recorded along with SATs and ACTs in college data files.

In the chart below, note the higher proportions of students from parochial and independent schools who chose to be Non-Submitters. It is perhaps a bit counterintuitive, given the proportional preference of underrepresented minorities and first-generation students to be Non-Submitters. But recognizing the average ratios of school counselor-to-students in these three types of high schools, we might imagine that students in public schools are getting less on-point advice about how to use a TOP to their advantage. Jerome Lucido comments on this issue in his recent essay on optional testing, positing that the gaps in TOP use by high school type may reflect wildly uneven college counseling resources (Lucido, 2018).

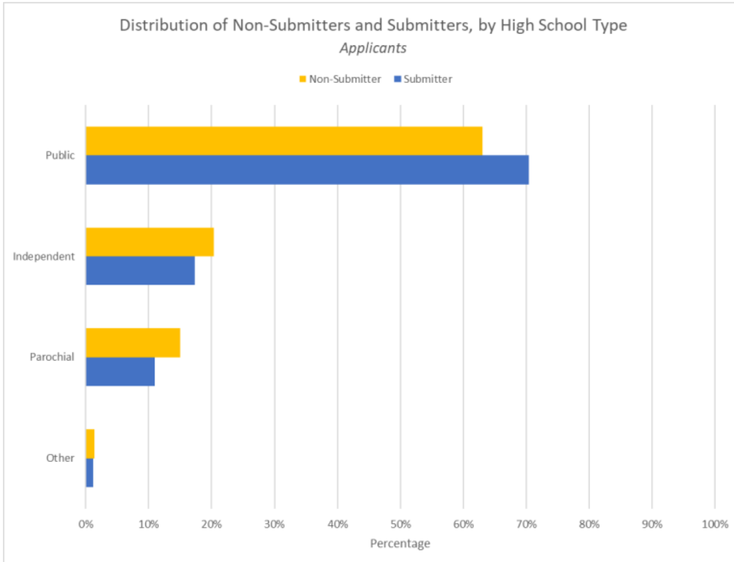


Figure 22. Distribution of Non-Submitters and Submitters by High School Type. **Pooled Student Data.** Seventeen institutions represented with reliable Submitter and Non-Submitter data on High School Type. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
17 Institutions **|** 335,904 Records
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $10,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-80% (43% M)

We are not surprised to see TOPs also being steadily used by students with excellent access to college counseling. It is a descriptive rather than judgmental comment that some Non-Submitters appear to be "accurately playing the corners" in their college applications, helped by advice from college counselors. These students largely do not have high testing in their favor, but they will have an accumulation of other attributes, starting with solid to spectacular transcripts, but often including evaluations for athletics, the arts, leadership, legacy status, geography, language or cultural backgrounds, and service commitments. In this respect, as in so many others, this policy turns out to be a helpful tool for complex and varied pools of students.

Because standardized tests often present special challenges for students with Learning Differences, both this study and the 2014 study attempted to ascertain whether LD students gravitated toward being Non-Submitters. Most of our institutions did not systematically collect this information during the application process, and even after students were enrolled, it was not systematically stored as retrievable data. We were, however, able to gather a small pool of information from nine institutions in the study.

As in "Defining Promise," we found LD students represented a higher portion of the Non-Submitters than the Submitters (7% versus 4%). However, the pool of data is limited, so can only suggest broader trends. As with other facets of TOP efforts, LD student access is a research project with potentially very high rewards, waiting to be done.

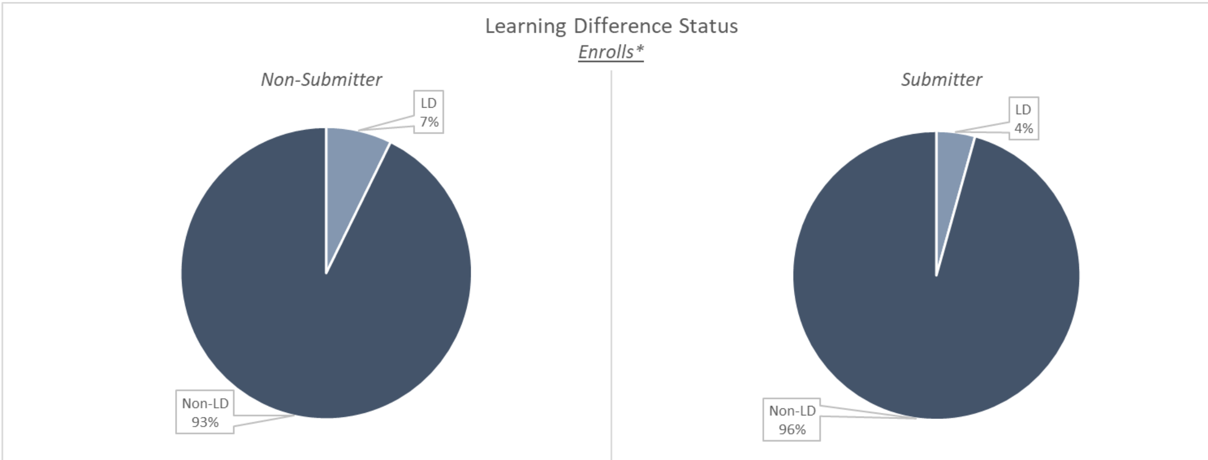


Figure 23. Percentage of Enrolled Students with Learning Differences (LD), Non-Submitter vs. Submitter Comparison. **Pooled Student Data.** Note that LD identification from institutions was provided at the enrolled student level. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
9 Institutions **|** 19,018 Records (of which 972 are LD)
Enrollment = 1,500-20,000 (2,000 M) **|** Endowment per FTE = $10,000-$800,000 ($40,000 M) **|** Admit Rate = 15%-65% (40% M)

We also noted equivalent proportions of athletes choosing to be Non-Submitters and Submitters which, in the aggregate, we suspect may be a cross-current of opposites. While Division I athletes are required by NCAA regulations to submit testing as part of their applications, a good many of the institutions in the study have teams at the D-III levels and their coaches have demonstrated a proclivity to actively encourage lower-testing athletic recruits to apply as Non-Submitters.

Similarly, we found equivalent rates of Early Decision or Early Action use among Submitters and Non-Submitters of testing, both at roughly a quarter of the applicants. So these "Early" plans seem to have equal appeal to either group.

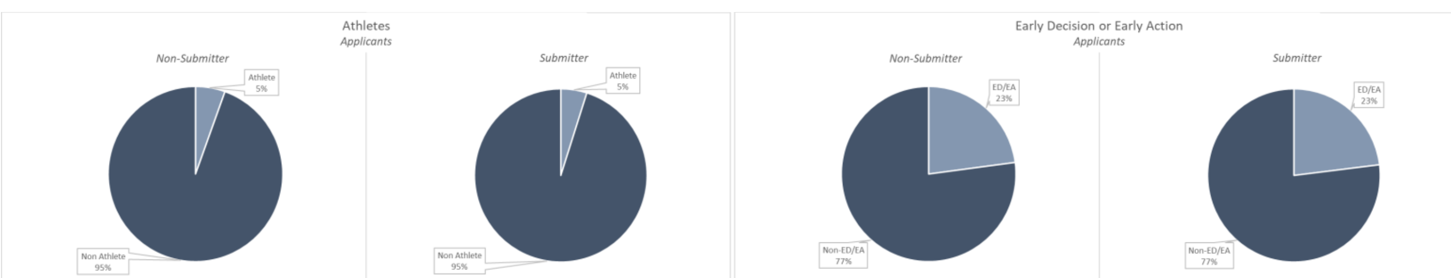


Figure 24. Percentage of Athletes and Early Decision/Early Action Applicants, Non-Submitter vs. Submitter Comparison. **Pooled Student Data.** Each set of charts represents data from a subset of institutions that provided reliable data on Athletes, and institutions that offered either an ED or EA program. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
Athlete: 16 Institutions **|** 261,047 Records
ED/EA: 22 Institutions **|** 437,318 Records
Enrollment = 1,500-10,000 (2,500 M) **|** Endowment per FTE = $10,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-80% (40% M)

An exploration of the use of TOP based on student home geography revealed a higher share of Non-Submitters in this study from the Middle Atlantic and New England states, but this may be partially explained by the distribution of TOP institutions. While our study includes institutions from 14 states, 20 of our 28 institutions are located in New England or the Mid-Atlantic. And of the 21 that provided data for this geographical comparison, 16 were in those two regions. There are higher concentrations of TOP colleges and universities on the East Coast and Non-Submitters appear to be somewhat more likely to apply regionally than nationally. According to FairTest's admittedly inclusive listings, there are now TOP institutions in 49 states, D.C. and most US territories. As TOPs are adopted by more institutions around the country, one would expect the geographic distribution of Non-Submitters to expand.

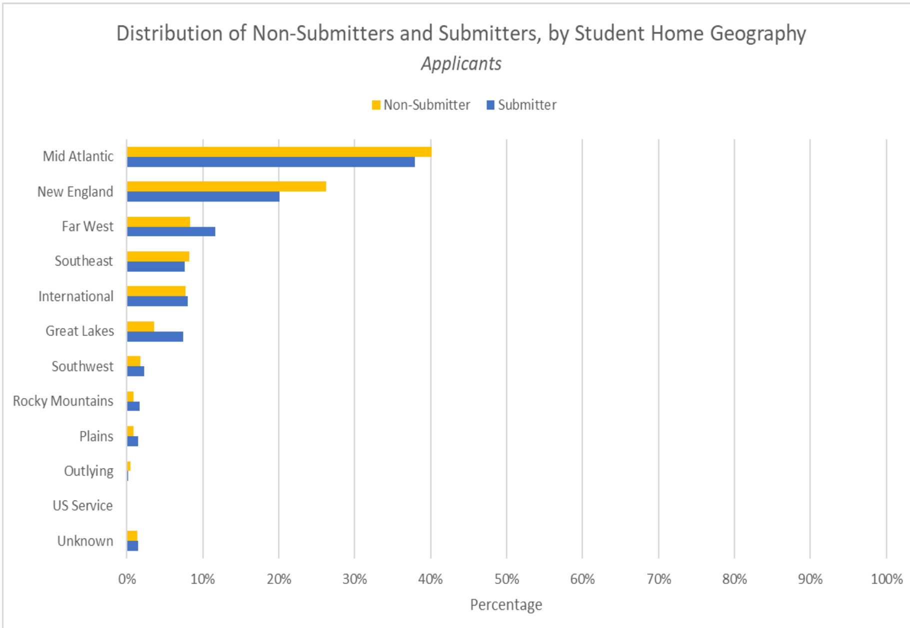


Figure 25. Distribution of Non-Submitters and Submitters by Student Home Geography. Pooled Student Data. IPEDS geography categories.
Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
21 Institutions **|** 343,621 Records
Enrollment = 1,500-10,000 (2,500 M) **|** Endowment per FTE = \$5,000-\$800,000 (\$70,000 M) **|** Admit Rate = 15%-90% (40% M)

How is "Expanded Diversity" represented among Non-Submitters?

Non-Submitters include higher proportions of students representing any combination of First-Generation, Pell, and URM students (i.e., each student counted only once): 42% of the enrolled Non-Submitters versus 32% of the Submitters. These results cast a new light on the attractiveness of a TOP for these underserved students and the opportunity for colleges to use this policy to attract and enroll more of them.

In recent decades, colleges and universities have put a great deal of effort into increasing the diversity of their campus communities. Correctly cited are ethical considerations to have colleges and universities serve social needs and offer access to underserved populations.

There are a multitude of characteristics currently identified in discussions of the "diversity" of a student population, and most center around the educational and developmental values associated with differences in perspective that result from differences in life experience. There is increasing concern and discussion about whether the use of standardized tests (specifically SAT and ACT) unnecessarily limits the admission of otherwise well-qualified students and tends to replicate the status quo in social class and opportunity in our American society.

> *…test scores appear to calcify differences based on class, race/ethnicity, and parental educational attainment. To come to some resolution, the Commission agrees that without confusing correlation with causation, admission offices must remain aware that test score differences persist among already under-served populations. Part of the public mission of colleges and universities is to ensure that differences that are not attributable to a student's ability to succeed academically at an institution are mitigated in the admission process. (NACAC Commission on the Use of Standardized Tests in Undergraduate Admission, 2008)*

We highlight three identifiable populations that have traditionally been under-represented in American higher education: First-Generation College Bound, students from lower SES backgrounds (using Pell Grant recipients as an easily-identifiable proxy), and students from racial and ethnic groups that have traditionally been underrepresented in college populations (URM). Often these are discussed as three distinct populations, failing to account for the overlap or intersectionality of the three. Thus, we offer the construct of "Expanded Diversity" in which we have included any student identified with any of these three groups.

Although we had a limited pool of seven institutions that were able to provide comprehensive data on URM, Pell recipients, and First-Generation students at the admit stage as well as at the enrolled stage, we offer the charts below to provide a visual representation of the richly more diverse opportunity presented by the pools of Non-Submitters admitted by these seven TOP institutions.

The first pie chart compares the Non-Submitters and Submitters that were offered admission by these seven institutions, and the second illustrates the students who actually enrolled from the same cohorts. In the admitted and the enrolled populations, just under **40% of Non-Submitters identify with one or a combination of these categories**, as compared to roughly **a quarter of Submitters**.

While we've already shown that each of the three subgroups (URM, First Gen, and Pell recipients) are more strongly represented among Non-Submitters, we draw your attention here to the sizable difference in representation of those students who are representatives of a **combination of all three** of those groups -- 9% of the enrolled Non-Submitters versus 3% of the enrolled Submitters. This serves to emphasize the attractiveness of a TOP for these underserved students and the opportunity for colleges to use this policy to attract and enroll more of them.

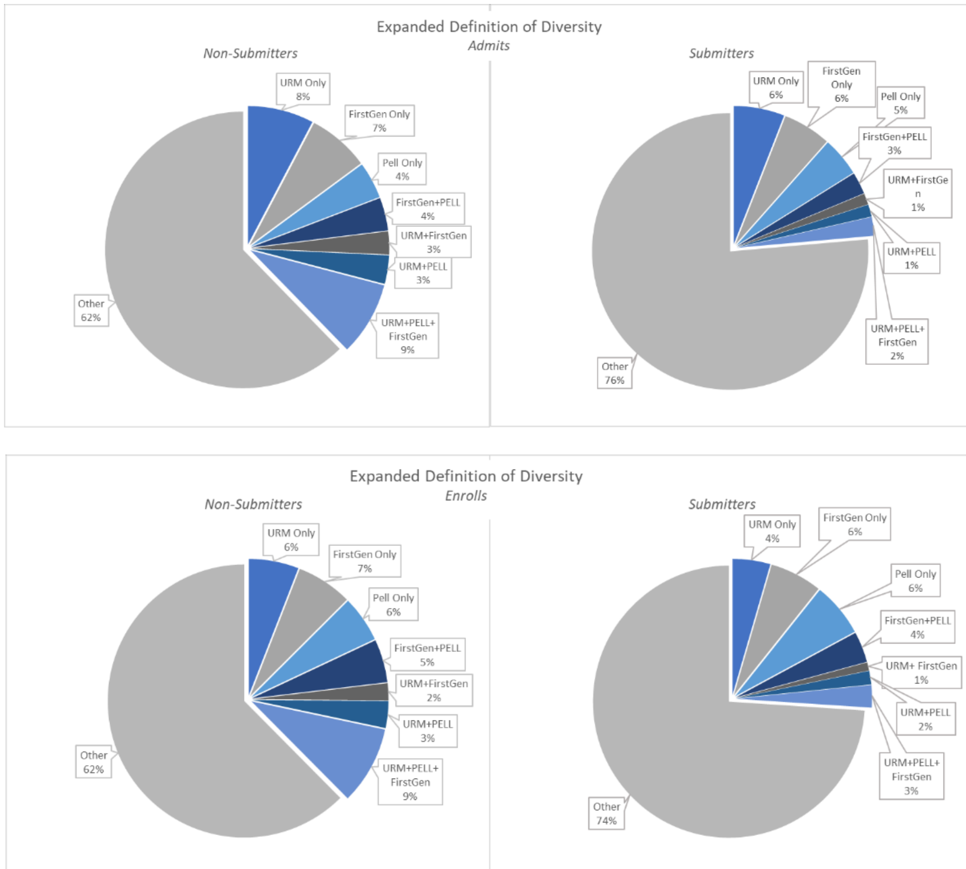


Figure 26. Defining Diversity Percentages at the Admit and Enroll Stages, Non-Submitter vs. Submitter Comparison. Pooled Student Data.
Enrolled students at the 7 public and private institutions that submitted reliable URM, First-Generation to College, and Pell-Recipients. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.
Admits: 7 institutions **|** 39,047 Records (of which 10,262 are "Expanded Diversity" Records)
Enrolls 7 institutions **|** 9,755 Records (of which 2,788 are "Expanded Diversity" Records)
Enrollment = 2,000-4,000 (2,500 M) **|** Endowment per FTE = $60,000-$500,000 ($150,000 M) **|** Admit Rate = 20%-70% (40% M)

Although only seven institutions provided all of the requisite data at the admit stage, there were 21 institutions in our study that provided reliable data on all three of these groups at the enrolled stage, so we've included the additional 14 institutions in the following representation. In this larger pool of enrolled students, 42% of Non-Submitters identify with one or a combination of these categories, as compared to about a third (32%) of Submitters.

And, though not represented in this chart, the differences between are even more pronounced at the individual institutional level. There, we see that Non-Submitters hold higher proportions of *total* Expanded Diversity, with the median institution at 43% for Non-Submitters vs 27% for Submitters. We also see that Non-Submitters comprise larger proportions of students who identify with more than one underrepresented group.[10] Amongst these institutions, the median is 19% for Non-Submitters vs 9% for Submitters.

[10] Combined total proportion of students who identify as URM+FirstGen, URM+Pell, FirstGen+Pell and URM+FirstGen+Pell

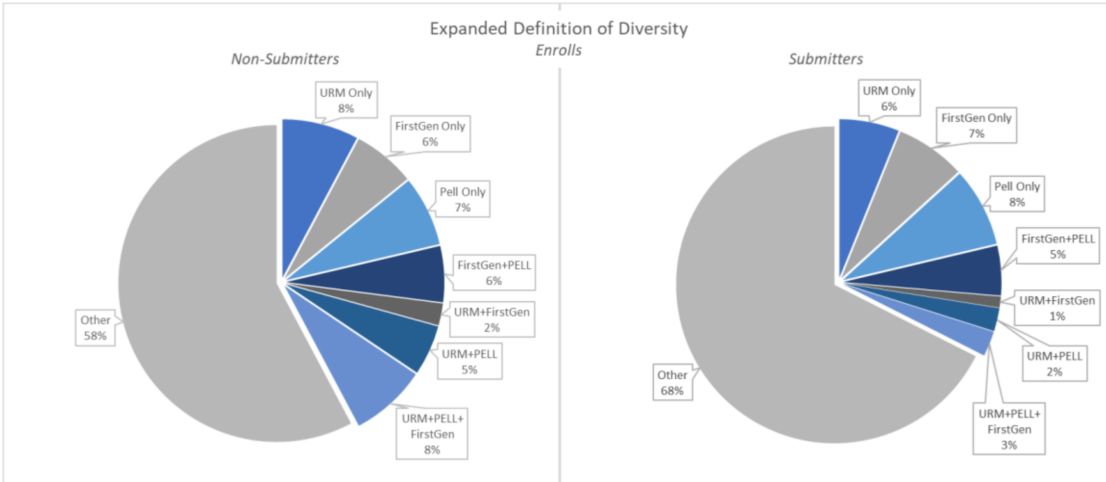


Figure 27. Expanded Diversity Enroll Percentages, Non-Submitter vs. Submitter Comparison. Pooled Student Data. Enrolled students at the 21 public and private institutions that submitted reliable URM, First-Generation to College, and Pell-Recipients. Exclusions: Two institutions with Academic Threshold policies, one that did not have Non-Submitter proportions.

21 institutions **|** 36,034 Records (of which 12,579 are "Expanded Diversity")

Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $7,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (40% M)

ACADEMIC OUTCOMES

Does adopting a Test-Optional Policy lower the academic quality of the enrolled student body?

In a word, no. Applicant average High school grades and SAT scores increased from pre-policy to post-policy at all but one of our institutions.

We sought to answer this question by comparing the average high school GPA and SAT scores from “pre” and the “post” policy years. While we could identify no reasonable way to compare these characteristics for TRP Peer institutions, and, in the context of rising grade inflation, we cannot confidently attribute these increases to the adoption of a TOP, we thought it worth sharing the experience of these 13 TOP institutions (those with cohorts from immediately preceding and following adoption of a TOP). While these institutions experienced only modest gains, all but one of the institutions experienced an increase in the average HSGPA of their applicants and again all but one saw an increase in their average SAT scores (though presumably the latter may be the result of some of the applicants with lower SAT scores choosing to apply as Non-Submitters).

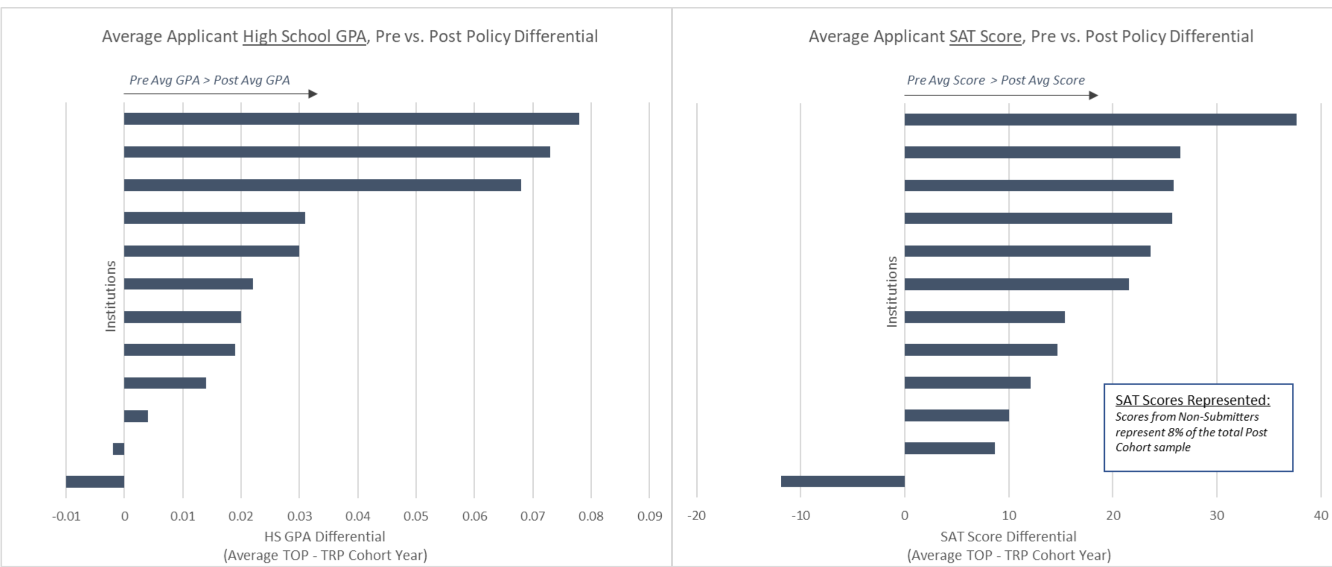


Figure 28. Average Applicant High School GPA and SAT Score Differentials: Pre-Policy vs. Post-Policy. The institutions represented include thirteen public and private TOP institutions in the study that provided reliable HSGPA and test scores for cohorts immediately preceding and immediately following their adoption of a TOP.
Exclusions: Two institutions with Academic Threshold policies and four early adopters.
13 Institutions **|** HS GPA = 536,011 Records **|** SAT = 560,016 Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$250,000 ($60,000 M) **|** Admit Rate = 20%-80% (50% M)

How do Non-Submitters and Submitters compare academically at each stage of the funnel?

In comparison with Submitters, the Non-Submitters showed slightly lower high school grades (-0.12), academic ratings, and college first year GPAs (-0.17). And, the Non-Submitters for whom we had scores had a significantly lower SAT average score. Once enrolled, Non-Submitters were less likely to designate a STEM major than Submitters.

With academic data pooled from virtually all of our participating institutions, the table below illustrates that the Non-Submitters had HSGPAs that were marginally lower at all three stages (applicant -0.12, admit -0.09, enroll -0.05) and, for those with SAT scores (26%, 31%, 35%, respectively, of Non-Submitters), they were lower than the Submitters by just under 200 points (185, 192, and 168 respectively). The only college statistic available at all 25 of these institutions was the FYGPA and here we see that the Non-Submitters lagged behind the Submitters by 0.17, consistent with the difference in their HSGPAs.

Academic Credentials at Each Stage of the Funnel

25 Public and Private Institutions

		Non-Submitters	Submitters
APPLICANT	*Applicant n*	*100362*	*378646*
	High School GPA	3.40	3.52
	Academic Rating	5.61	6.06
	SAT	1041	1226
ADMIT	*Admit n*	*36843*	*192588*
	High School GPA	3.53	3.62
	Academic Rating	6.69	6.89
	SAT	1095	1253
ENROLL	*Enrolled n*	*10878*	*38318*
	High School GPA	3.52	3.57
	Academic Rating	6.19	6.52
	SAT	1082	1226
	First Year GPA	3.07	3.24

Comparison of Means (Cohen's d)	
No Difference	< 0.1
Small Difference	0.1 - 0.3
Medium Difference	0.3 - 0.5
Large Difference	> 0.5

SAT Score Representation for Non-Submitters:
Applicants: 26% of Non-Submitters had a score on record
Admits: 31% of Non-Submitters had a score on record
Enrolls: 35% of Non-Submitters had a score on record

Figure 29. Academic Profile of Non-Submitters vs. Submitters at Each Stage of the Funnel. Pooled Student Data.
Academic Rating: 20 of the 25 institutions in this analysis submitted Academic Ratings. They each submitted their own scales, but for comparison purposes we converted all to a 10-point scale, where 10 is the highest rating.
Exclusions: Two with Academic Threshold policies and one that did not have Non-Submitter proportions.
25 Institutions **|** 479,008 Records
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $6,000-$800,000 ($70,000 M) **|** Admit Rate = 15%-90% (40% M)

> *"From faculty anecdotal feedback, this may be the best class we have seen, in terms of student curiosity, involvement, etc. The Admissions staff is pleased by the access and inclusion we see in the class."*
>
> *Dean from small private university*

Next, we considered differences between Non-Submitters and Submitters in choice of majors. Consistent with earlier studies, Non-Submitters were more likely than Submitters to declare majors in: Humanities and Liberal Arts, Social Science, and Psychology and Social Work. And Submitters were more likely than Non-Submitters to select Business, Biology and Life Science, Computers and Mathematics, and Education.

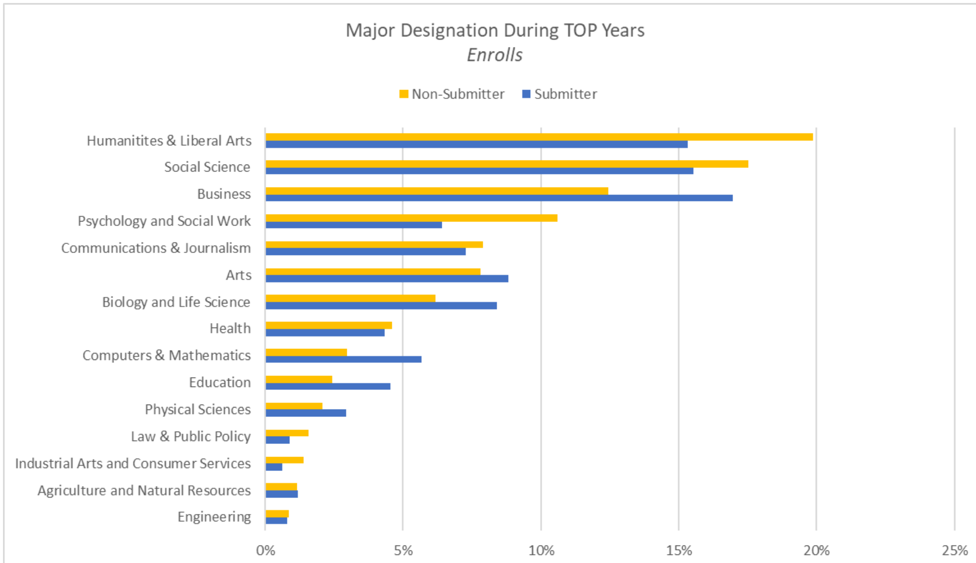


Figure 30. Academic Outcomes: Major Designation. Pooled Student Data. Enrolled students at 20 public and private institutions with at least two years of a TOP policy in place in order to have declared majors (CIP codes 2010). CIP codes were converted to major categories designated by the Center for Education and the Workforce, Georgetown.
Exclusions: Two with Academic Threshold policies and one that did not have Non-Submitter proportions.
20 institutions **|** 31,692 Records
Enrollment = 1,500-10,000 (2,500 M) **|** Endowment per FTE = $6,000-$800,000 ($60,000 M) **|** Admit Rate = 15%-90% (43% M)

The chart above illustrates that Submitters were more likely to have chosen STEM disciplines than Non-Submitters, 23% vs 18%, respectively. So, are Non-Submitters selecting higher-grading disciplines, thereby lifting the overall average GPA for the group?

To answer this question, we identified STEM majors and Non-STEM majors to assess the impact on the Cumulative GPAs of Submitters and Non-Submitters. The chart below, indicates that for this group of 31,000 students there is not a dramatic difference between the GPAs achieved by those majoring in STEM fields versus those in non-STEM fields. So, for this sample, the choice of major does not appear to have had a significant influence on the Cumulative GPAs of either group.

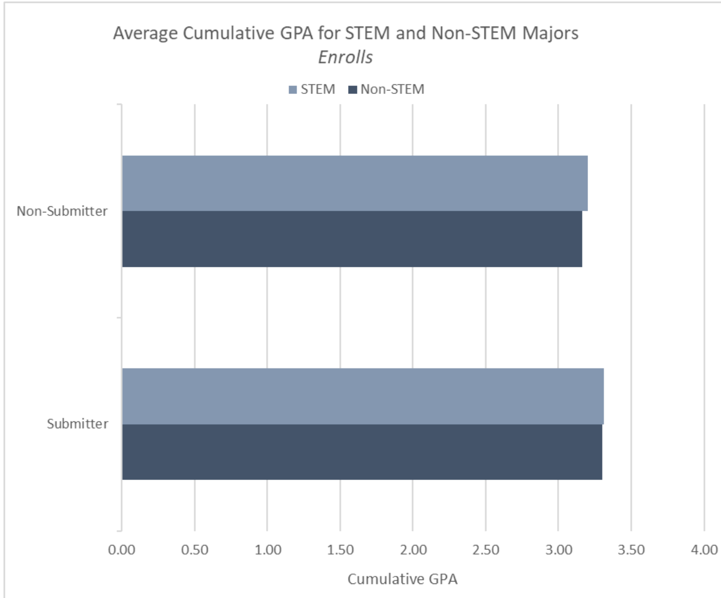


Figure 31. Average Cumulative GPA of STEM Majors and Non-STEM Majors, Non-Submitter vs. Submitter Comparison. Pooled Student Data.
Enrolled students at **20** public and private institutions with at least two years of a TOP policy in place in order to have declared majors (CIP codes 2010). STEM majors were identified using Department of Education's Classification of Instructional Programs[11]
Exclusions: Academic Threshold, Institutions with very recent TOP Policies
20 institutions **|** 31,692 Records
Enrollment = 1,500-10,000 (2,500 M) **|** Endowment per FTE = $6,000-$800,000 ($60,000 M) **|** Admit Rate = 15%-90% (43% M)

How do the academic outcomes compare with those of "Defining Promise"?

In short, Non-Submitters performed quite well in both studies. While their college First Year GPAs and Cumulative GPAs were marginally lower than Submitters, both this study and the 2014 study confirmed they graduated at rates equivalent to or slightly above Submitters.

Reflected below are parallel data points from the 2014 "Defining Promise" study, which included only enrolled students. To allow us to better compare the outcomes in the two studies, we included only the 13 institutions from "Defining Access" that provided 4- **and** 5-year graduation rates. Because this chart includes only about half of the institutions from Fig 31 above, for reference, we have included the applicant pool data for these students. The HSGPAs are slightly higher for this group than the larger group above (0.13 for Non-Submitters and 0.09 for Submitters), as are the SATs.

A major finding of the earlier study was that were no significant differences in the First Year GPA, Cumulative GPA or graduation rates between Submitters and Non-Submitters. We analyzed the 2018 data in the same manner and found that the results were strikingly similar.

[11] STEM has been defined as the Department of Education's Classification of Instructional Programs taxonomy within the two-digit CIP series containing engineering, biological sciences, mathematics, and physical sciences, or a related field. These fields represent research, innovation, or development of new technologies using engineering, mathematics, computer science, or natural sciences (including physical, biological, and agricultural sciences).

	Defining Promise, 2014 (20 Private Institutions)		Defining Access, 2018 (13* Public and Private Institutions)	
	Non-Submitters	Submitters	Non-Submitters	Submitters
Applicant n			37904	110913
Applicant High School GPA			3.48	3.60
Applicant Academic Rating			5.98	6.90
Applicant SAT			1124	1243
Enrolled n	2427	4064	4282	12846
High School GPA	3.50	3.52	3.63	3.64
Academic Rating	6.86	6.78	6.05	6.67
SAT	1085	1231	1108	1234
First Year GPA	3.02	3.12	3.03	3.21
Cumulative GPA (Non-Grad)	2.74	2.83	2.70	2.79
Cumulative GPA (Graduated)	3.24	3.33	3.23	3.40
Graduation Rate 4-Year	78.4%	76.8%	79.0%	76.4%
Graduation Rate 5-Year	78.8%	77.1%	82.8%	82.0%

Comparison of Means (Cohen's d)	
No Difference	< 0.1
Small Difference	0.1 - 0.3
Medium Difference	0.3 - 0.5
Large Difference	> 0.5

Comparison of Proportions (Chi-Square)	
No Significant Difference	p > 0.05
Significant Difference	p < 0.05

SAT Scores for Defining Promise:
Enrolls: 41% of Non-Submitters have test scores on record.

SAT Scores Defining Access:
Applicants: n= 6681 Non-Submitter Test Scores, or 18% of applicants.
Enrolls: n = 1295 Non-Submitter Test Scores, or 30% of enrolls.

Figure 32. Two Study Comparison[12]: Academic Profile of Non-Submitters vs. Submitters. Pooled Student Data.
Defining Promise: 20 private institutions (no minority serving or arts institutions represented) that submitted 4-**and** 5-year graduation rates (students entering cohort 2008 for 4-year; students entering 2007 for 5-year)[13]
Defining Access: 13 (*12 private and 1 public) institutions that submitted 4-and 5-year graduation rates (cohorts entering 2012 for 4-year **and** 2011 for 5-year). Exclusions from both studies: Academic Threshold policies
Academic Rating: All institutions submitted their respective scales, but for comparison purposes we converted all to a 10 point scale, where 10 is the highest rating.
Enrollment = 1,500-5,000 (2,000 M) **|** Endowment per FTE = $6,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (40% M)

The HSGPAs were equivalent between Submitters and Non-Submitters, but there was a significant difference in the SAT scores.[14] The Non-Submitter average was lower by 146 points in the 2014 study and 126 points in the current study.

In college, the Non-Submitters started their college careers with FYGPAs that were lower than Submitters by 0.10 to 0.18, a GPA pattern that persisted through to graduation. In both studies, both Non-Submitters and Submitters that graduated earned college Cumulative GPAs over 3.20, and those that did not graduate (from that institution) posted Cumulative GPAs in the 2.7-2.8 range. But in spite of having slightly lower GPAs, both studies found Non-Submitters graduating at marginally higher rates than Submitters.

Because the above data is pooled data from the included colleges, we also examined the comparative Submitter and Non-Submitter graduation rates *at the institution level*. While the bulk of the colleges reported the same result as the pooled averages above—little difference between Submitter and Non-Submitter graduation rates—there were four private institutions that experienced significantly lower rates among Non-Submitters. It is important to note that two of these four were identified earlier in the report as having also experienced lower application growth, lower URM

[12] Six institutions in this subset participated in both studies. However, there is ***no cohort overlap***. Each study represents a different set of students.
[13] Please note that based on the way the data was submitted, the 4 year and the 5 year graduation rates reflect two separate cohorts of students from the same set of institutions.
[14] The scores for enrolled Non-Submitters represent only 41% and 30% of the Non-Submitters, respectively, though it can be argued that it is unlikely that Non-Submitters with higher scores would disproportionately choose not to share them. As more institutions have become comfortable with TOP, fewer of them seem to feel compelled to collect test scores from matriculating students to allow them to conduct research – hence the lower proportion of test scores available for Non-Submitters in 2018.

enrollment growth, and lower Pell enrollment growth compared to their matched TRP Peers. Looking back into the institutional analyses and interview notes done for each institution in our study, we found some common patterns at these schools:

- Very high levels of self-help in financial aid packages for High-Need students.
- Variability in graduation rates from one cohort to the next, with one of the two cohort years showing disparately poor graduation rates for Non-Submitters.
- Larger than average gaps in HSGPA (.3) between Non-Submitters and Submitters.

In the chart below we've isolated the graduation rates of students who were identified as part of the Expanded Diversity group (URM, First Gen, and/or Pell). Here we note that Non-Submitters graduated at a rate 6%-7% higher than Submitters. The students represented in the "Other" category show more equivalence between Submitters and Non-Submitters.

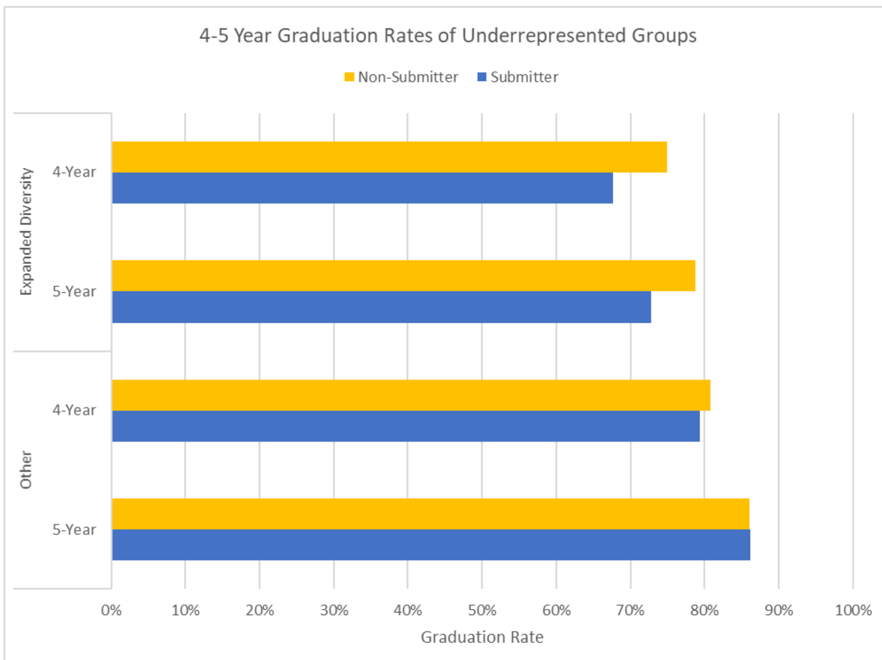


Figure 33. Four to Five Year Graduation Rates, Expanded Diversity: Non-Submitter vs. Submitter Comparison. Pooled Student Data. Enrolled students at public and private institutions with mature TOP policies, as to have 4 **or** 5 year graduation rates, and reliable data on each underrepresented group.
13 Institutions **|** Expanded Diversity: 3814 Records **|** Other: 8696 Records
Enrollment = 1,500-7,000 (2,000 M) **|** Endowment per FTE = $6,000-$800,000 ($125,000 M) **|** Admit Rate = 15%-90% (43% M)

Under TOP, how well did the High School GPA correlate with college success?

High school GPA had a strong positive linear correlation with college cumulative GPA, and it had a stronger relationship with both GPA and graduation rate than the SAT/ACT score. The test scores showed stronger correlation with Submitter outcomes than Non-Submitter outcomes. They also had very strong correlation with EFC.

We pitted the available admission academic markers—high school GPA and SAT scores—against our college success markers, college GPAs and graduation rates. The scatterplots below have been constructed to illustrate these relationships side by side, using averaged percentiles[15] of student records containing both sets of data points. To be clear, charts with SAT scores include <u>all</u> the students for whom we have a test score, but while the chart accurately portrays the information for the two groups, the Non-Submitter group represents only about a third of all the Non-Submitters. We therefore offer these comparisons not as definitive but simply observational.

Intuition would suggest that those Non-Submitters with higher scores would be more likely to share them than those with lower scores, but there is no way to validate that hypothesis. At the eight institutions in the study that submitted test scores for more than 50% of Non-Submitters, however, we found the patterns to be identical.

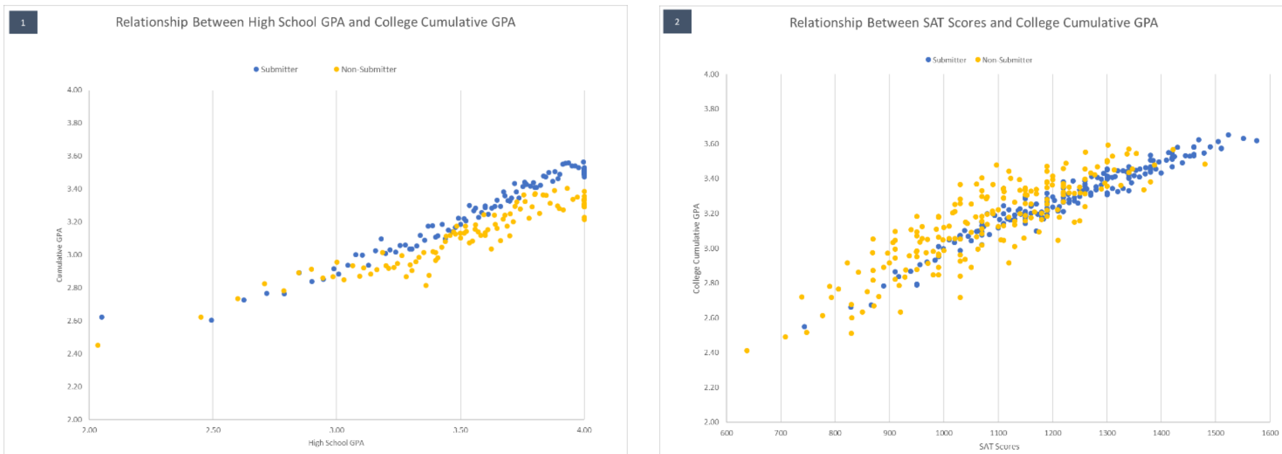


Figure 34. Relationships Between HSGPA, SAT and *<u>Cumulative GPA</u>* Under TOP Cohort Years, Submitter vs. Non-Submitter Comparisons.
1. HSGPA vs. CumGPA - Enrolled students at 20 public and private institutions. 41,320 Records
2. SAT vs. CumGPA - Enrolled students at 20 public and private institutions. 36,378 Records

Exclusions: Academic Threshold policies and very recent policy adopters.
Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($50,000 M) **|** Admit Rate = 15%-80% (43% M)

In Fig 34, Chart 1, there is a positive linear relationship between HSGPA and college Cum GPA, with a clear visual pattern of Submitters having modestly higher college Cum GPAs than their Non-Submitter classmates who had similar HSGPAs. This difference averages out to approximately the 0.17 of a GPA point that we cited earlier in this study, the difference between 3.23 and 3.40.

In the SAT to Cum GPA comparison (Chart 2), the relationship is also positive and linear. However, Non-Submitter score averages ranged lower than Submitters and were more scattered in their college GPAs at any given test score, and were more likely to have achieved higher college GPAs than Submitters with the same test scores. When making any interpretations of this particular chart, the reader must remember that the test scores here represent only 33% of the Non-Submitters.

[15] The points in the scatterplots were calculated using Analytics software, by Rapid Insights. The points represent records that contained both academic measures. The data was calculated into percentiles, then averaged for each cluster. For example, in the case of HSGPA vs CumGPA, each Submitter point is represented by 300 data points or 3% of the data.

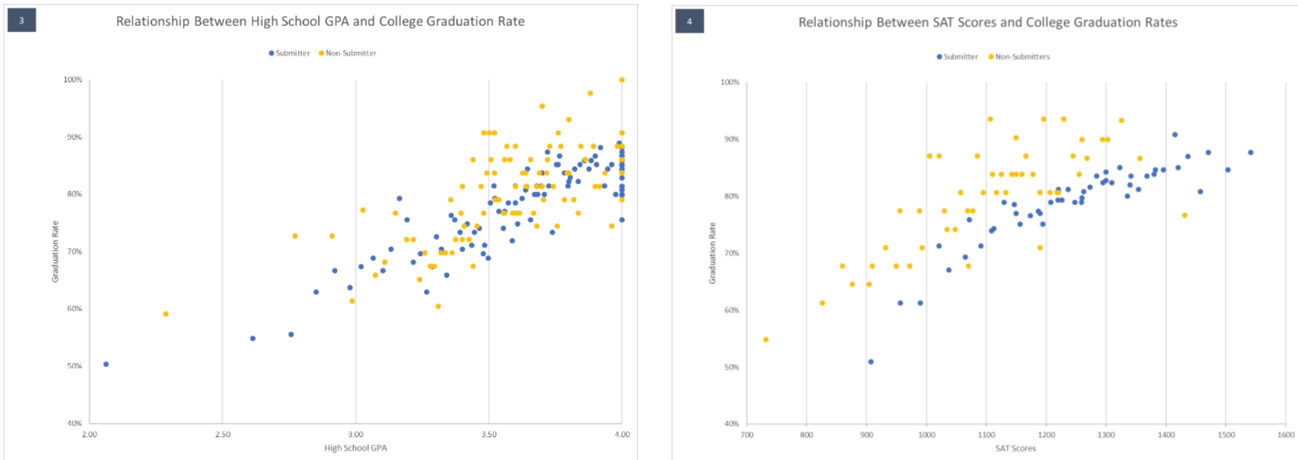


Figure 35. Relationships Between HSGPA, SAT and *Graduation Rate* Under TOP Cohort Years, Submitter vs. Non-Submitter Comparisons.
3. HSGPA vs. Grad Status – Enrolled students 17 public and private institutions with 4 or 5 year graduation rates. 17,798 Records
4. SAT vs. Grad Status – Enrolled students 17 public and private institutions with 4 or 5 year graduation rates graduation rates. 14,593 Records
Exclusions: Academic Threshold policies and recent policy adopters.
Enrollment = 1,500-7,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-80% (40% M)

In Fig 35, Chart 3 we have plotted HSGPA against graduation rates. (N.B. All graduation rates in this study represent only same-school graduation rates.) While the relationship is not as strong as in Fig 34, Chart 1, as would be expected there is a general upward trend correlating higher high school grades with higher college graduation rates. Chart 4, plotting SAT scores against graduation rates, reveals a modestly different pattern. Submitters show a tighter, more linear correlation between SAT scores and graduation rates, but a majority of the Non-Submitters graduated at higher rates than Submitters with comparable SAT scores. Note that the number of Non-Submitter test scores is less than 1,600, so this is a small sample size, and represents only about 31% of the Non-Submitters at these institutions. Also note that virtually none of the Non-Submitters have scores in the highest ranges.

Chart 5 and 6 below in Fig 36 are somewhat limited because they include only students who applied for financial aid, and thereby miss the upper portions of the income ladder. Nonetheless, Chart 5 illustrates the ***lack of*** relationship between high school grades and EFC. But the chart reveals what all of the data to date has shown: that Non-Submitters will tend to be from lower-income families, but many of them have fine high school records.

Chart 6 is in stark contrast. We see a strong correlation between SAT scores and family affluence (in this case using EFC as a proxy for other family financial data). The relationship with Non-Submitter scores is mildly more diffuse, which actually aligns well with these students' sense that the test scores do not represent them well[16].

Also note that the EFC ranges appear different between the two charts. This is due to the relative strength of correlation between GPA and EFC versus SAT and EFC. Each data point on the chart represents a cluster of students. In chart 5, any clustering has students with both high and low EFCs causing their means to trend toward the middle of the overall range (i.e., for any particular GPA, there are students with a wide range of EFCs). The opposite effect can be seen in chart 6, where the correlation is much stronger and upper end of the averages remain high representing a lack of lower EFC students achieving those higher SAT scores.

[16] Remember that the SAT and EFC charts represent test scores from approximately a third of the Non-Submitters, and also do not include the upper part of the income ladder, since it only captures students who filed a financial aid application.

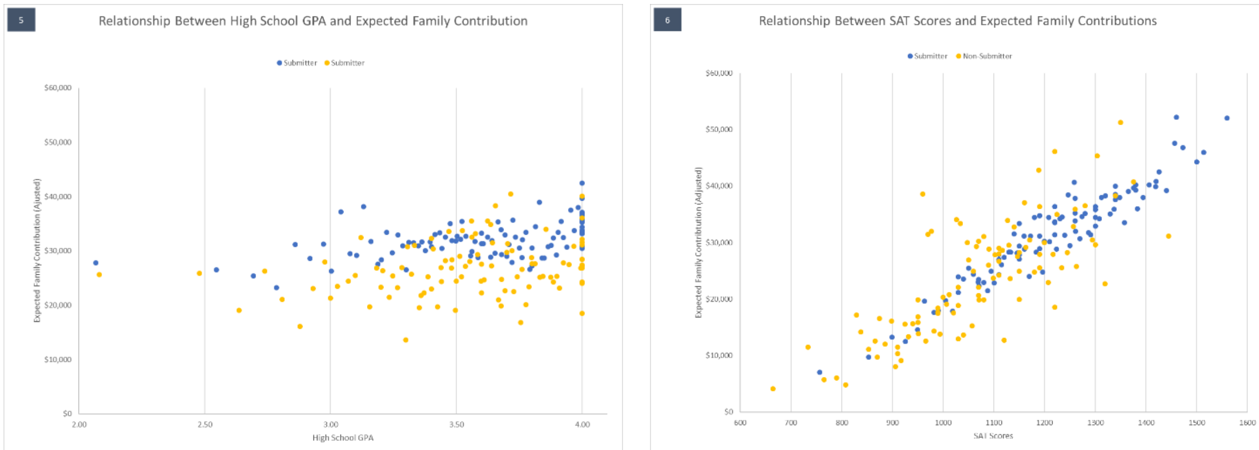


Figure 36. Relationships Between High School GPA, SAT and *__Expected Family Contribution__* Under TOP Cohort Years, Submitter vs. Non-Submitter Comparisons.

5. HSGPA vs. EFC – Enrolled students at 20 public and private institutions with reliable HSGPA and EFC data. 25,257 Records
6. SAT vs. EFC – Enrolled students at 20 public and private institutions with reliable SAT and EFC data. 21,333 Records

Exclusions: Academic Threshold policies.

Enrollment = 1,500-20,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($70,000 M) **|** Admit Rate = 15%-90% (43% M)

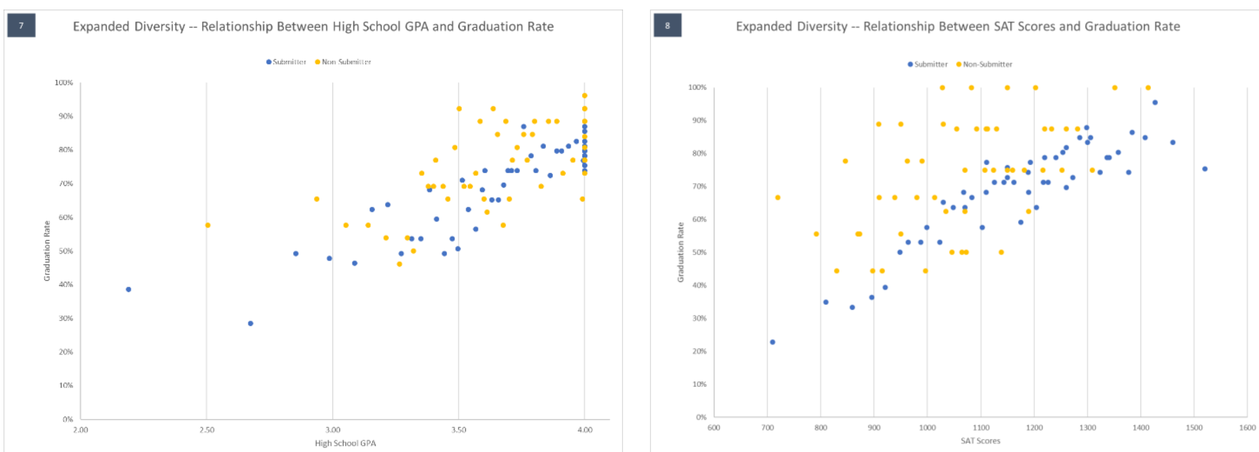


Figure 37. Focus on __Expanded Diversity__ -- Relationships Between High School GPA, SAT and *Graduation Rate* Among this Student Group, Submitter vs. Non-Submitter Comparisons.

7. HSGPA vs. Grad Status – Enrolled students at 13 institutions with reliable expanded diversity data, and 4 or 5 year graduation rates. 4,751 Records
8. SAT vs. Grad Status – Enrolled students at 13 institutions with reliable expanded diversity data, and 4 or 5 year graduation rates. 3,719 Records

Exclusions: Academic Threshold policies.

Enrollment = 1,500-7,000 (2,500 M) **|** Endowment per FTE = $5,000-$800,000 ($130,000 M) **|** Admit Rate = 15%-90% (43% M)

In Fig 37, charts 7 and 8, we have again used our construct of "Expanded Diversity" to explore, for this set of students, the correlations between graduation rates and either HSGPA or SAT scores. In Chart 7, graphing HSGPA against graduation rate, we see more or less equivalent positive relationships for Submitters and Non-Submitters, with a marginally tighter relationship for Submitters than for Non-Submitters, and with Non-Submitters more often clustered above the Submitters in graduation rate for students with a given HSGPA.

In chart 8, graphing SATs against graduation rate, the differences are more pronounced: a positive relationship for these underrepresented student (Expanded Diversity) Submitters, but a scattershot of graduation rates for Non-Submitters, mostly on the upper portion of the graph, suggesting that they were more promising students than their SATs would have suggested. (N.B. There are only 420 Non-Submitters represented in Chart 8, so this is observational, not conclusive.)

So, at least for this limited sample of Non-Submitters, high school GPA correlated more strongly than the SAT, with success in college, in terms of both college cumulative GPA and graduation rate. The SAT, while showing solid correlations with cumulative GPA for Submitters, continues to most strongly correlate with family affluence.

THE FINANCIAL SIDE OF THE POLICY

The following section includes some of the most complex analyses in the study. Herein, in addition to comparing differences between pre- and post-adoption cohorts, Non-Submitter and Submitter groups, aspects of diversity, and graduation rates, we have introduced information about the financial need and aid awards of students. Adding another level of interpretative complexity, we were not able to ascertain whether institutions were need-blind or need-aware at any (or all) points in the data of the study. Some (particularly more selective, more affluent) institutions utilize a need-aware admission policy to carefully regulate their commitment to financial aid. This, obviously, would find expression in the admit rates (particularly of high-need students) as well as in the aid offers made to students. Thus, what follows is a representation of the data we had available, without any effort to overlay these factors into our observations.

Additionally, institutions use different protocols in the awarding of aid to international students – some create an EFC, some award aid to international students but record EFCs only for domestic students, etc. -- so for the comparisons on aid, we have excluded all students identified as non-resident aliens (international students). By doing so, we hope the observations about need-based aid will be clearer.

Do TOPs place greater financial aid demands on the institution?

Our data limits drawing conclusions, but our findings suggest yes. High need students choose to be Non-Submitters at higher rates, and Gift Aid per Capita increased pre-to-post policy adoption.

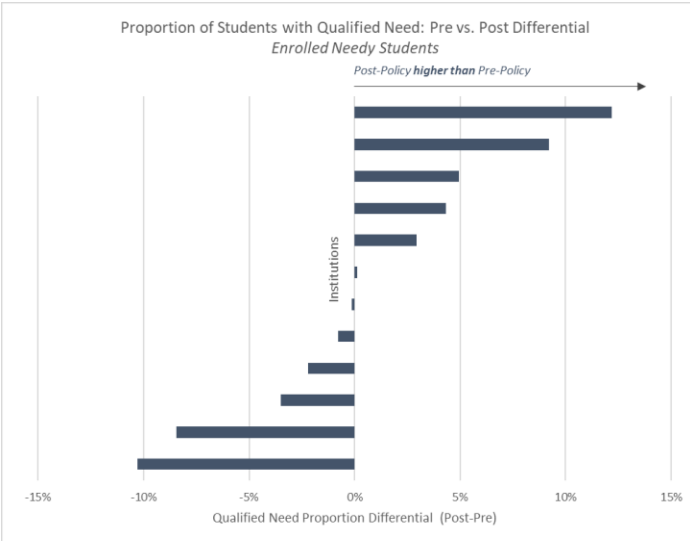
The chart below on the left compares the proportion of students with a qualified need[17] in the pre-policy and post-policy years for the 12 institutions[18] that provided data from the years immediately preceding and immediately following adoption of the policy. Among these 12 institutions, the changes in proportion of needy students were roughly evenly split, with five institutions experiencing an *increase* in the proportion of enrolled needy students, four experiencing a decrease, and three enrolling essentially the same proportion (within +/- 2%). As noted above, some of this may have been the result of need-aware admission policies.

However, the chart below on the right, shows that while the proportion of enrolled needy students did not necessarily increase with the adoption of a TOP, the average financial need[19] of the needy students did tend to increase at modest levels (corrected for inflation).

[17] The proportion of students who demonstrated a need for financial aid (Demonstrated Need > 0)

[18] Three of our institutions provided data for only post-adoption cohorts; there were several years of separation between the pre- and post-adoption cohorts for some of our other institutions; and still others were unable to provide reliable financial aid data for all their cohorts.

[19] Demonstrated Need was calculated as: Total Cost of Attendance, adjusted – EFC, adjusted.



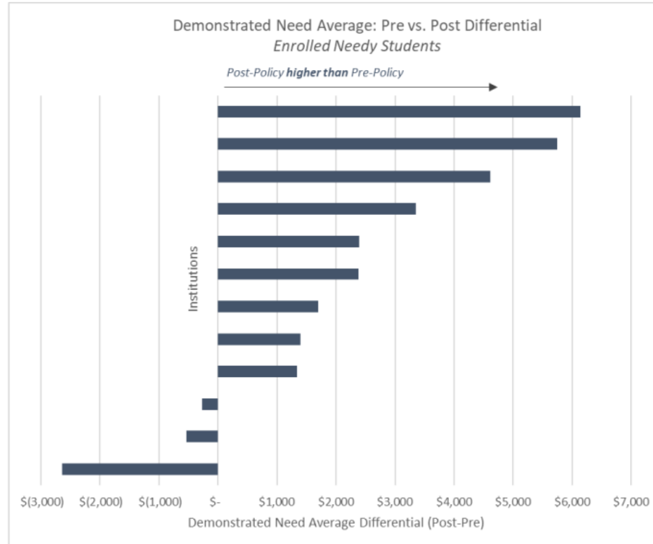


Figure 38. Qualified Need Proportions and Demonstrated Need Differentials: Pre-Policy vs. Post-Policy. Data from the 12 private and public institutions with EFC data submitted immediately pre- and immediately post-policy adoption.
Bars in LEFT and RIGHT charts do not align to reflect same institution. Note that all financial data has been adjusted to 2016-dollar standards.
Exclusions: Early adopters and Non-Resident Aliens
12 institutions **|** 36,912 Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$230,000 ($65,000 M) **|** Admit Rate = 20%-80% (53% M)

Expanding beyond the need of the enrolled students, we attempted to assess the actual financial impact on an institution by generating estimates of Gift Aid per Capita[20] for *all enrolled students*. Below we have compared the cohorts immediately preceding and following the policy adoption. We see that per capita costs rose during the TOP cohort years at all but one of these ten institutions. The smallest increase was $728 per capita, and the largest almost $4,000 per capita.

[20] Gift Aid per Capita was calculated as: Sum of Total Gift Aid-Adjusted / Total Enrollment.

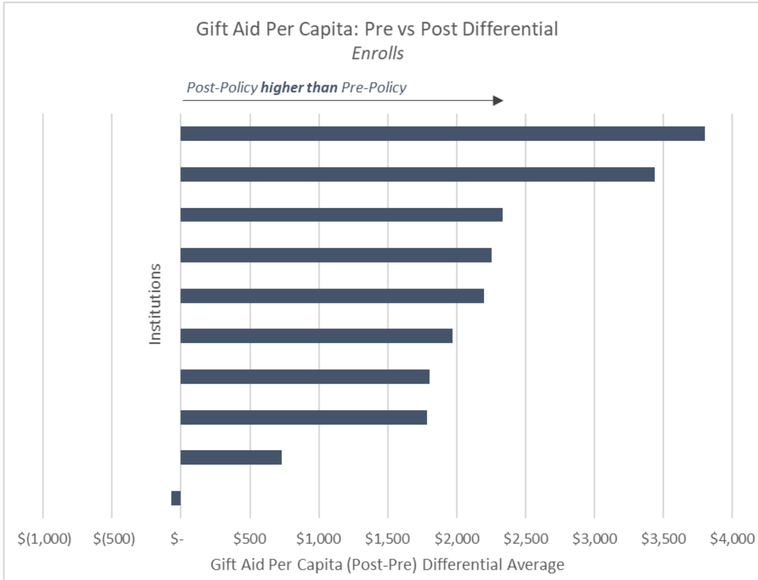
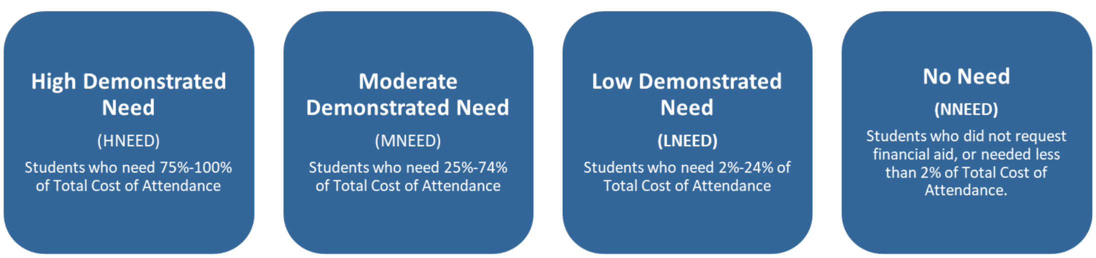


Figure 39. Gift Aid Per Capita Differential: Pre-Policy vs. Post-Policy. Data from the 10 private and public institutions with gift aid submitted immediately pre- and immediately post-policy adoption.
Exclusions: Early adopters and Non-Resident Aliens
Note that all financial data has been adjusted to 2016-dollar standards.
10 institutions **|** 56,564 Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$230,000 ($65,000 M) **|** Admit Rate = 20%-80% (53% M)

To what extent is this rise in investment a result of the policy? Although we cannot answer this question directly, we investigated the financial need of the incoming Non-Submitters for an indication. We divided students into the following segments to by subtracting the adjusted EFC from the Total Cost of Attendance. (A more detailed description of segments and this methodology can be found on page 75.)



Using these segments, the following Figure contrasts the pooled data for enrolled Non-Submitters and Submitters at the broader set of 21 institutions with reliable data at the enrolled stage. As might be expected, due to the attractiveness of the test-optional policy to lower SES students, the Non-Submitters had a higher proportion of High-Need students than Submitters (36% vs. 28%). However, on the other end of the financial spectrum, both Non-Submitters and Submitters had substantial proportions (34% and 38%) of No-Need students.

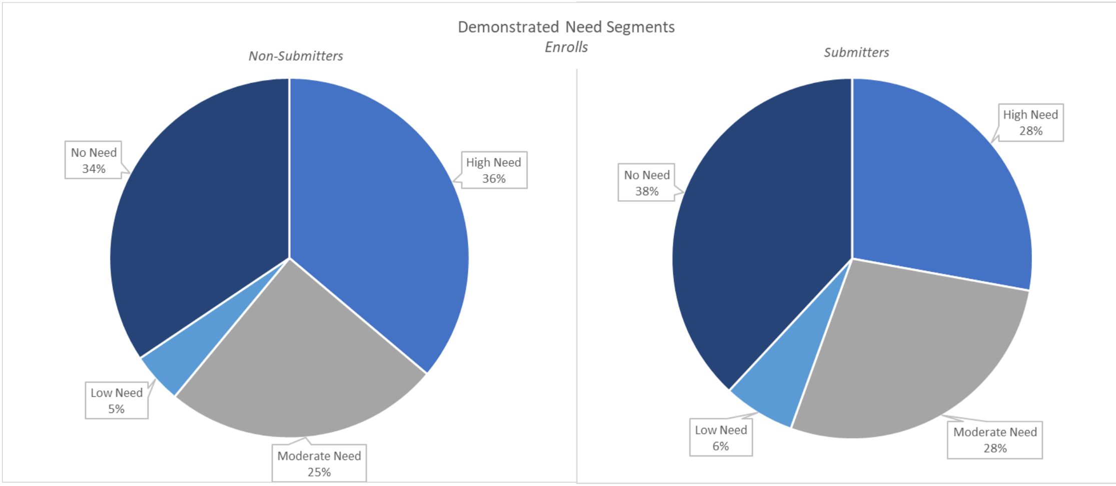


Figure 40. Demonstrated Need Segment Proportion Profiles: Non-Submitter vs. Submitter Comparison. Data from 21 private and public institutions with reliable financial aid data at the enroll stage of the funnel. Pooled Student Data.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
21 institutions **|** 34,305 Records
Enrollment = 1,500-20,000 (2,400 M) **|** Endowment per FTE = \$6,000-\$800,000 (\$52,000 M) **|** Admit Rate = 15%-90% (43% M)

The next chart provides an institution-by-institution comparison of the difference between the demonstrated need of the needy Submitters and Non-Submitters enrolling at the same 21 institutions. At every one of these institutions, the enrolled <u>needy</u> Non-Submitters had a higher average demonstrated need than the <u>needy</u> Submitters, with the median differential being approximately \$4,000.

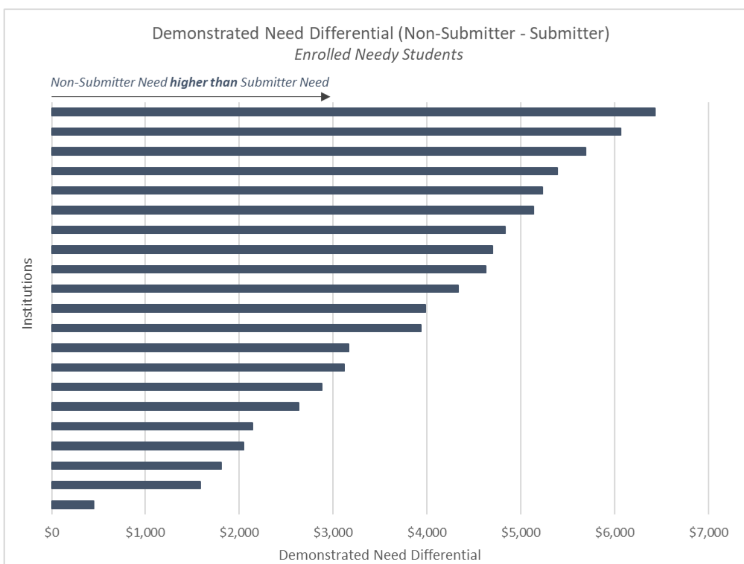


Figure 41. Average Enrolled Student Demonstrated Need Differential: Non-Submitter vs. Submitter. Private and public institutions with reliable EFC data on enrolled students with a qualified need.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
21 institutions **|** 21,564 Records
Enrollment = 1,500-20,000 (2,400 M) **|** Endowment per FTE = \$6,000-\$800,000 (\$52,000 M) **|** Admit Rate = 15%-90% (43% M)

Do institutions aid Non-Submitters differently than Submitters?

There were substantial differences in awarding strategies between institutions, with the majority having provided less generous (gift) aid packages to both their needy and no-need Non-Submitters than to their Submitters.

As described in detail in the Appendix on page 75, we approached this analysis by clustering student financial support in five "Need Met with Gift" Segments. Traditionally, an institution "meets need" with a combination of gift aid (grants and scholarships), student loans, and work-study. However in this study, we evaluated whether demonstrated need has been met solely with gift aid. Thus, in these clusters, many of the High-Need students (and even Moderate-Need students) will fall into our category of "Need Not Met with Gift." In some instances these students may have been "gapped," but the reader also should not conclude that students in this category didn't have their need met under the more traditional definition. It should also be noted that some institutions required test scores to be considered for scholarships which would tend to mildly skew some of these outcomes, particularly in the category of No-Need + Gift.

Using this framework, the following pair of charts uses pooled data and represents the "average" experience of a Submitter or Non-Submitter. It offers a comparison of the distributions at the admit stage and at the enrolled stage. At both stages of the funnel we see that Non-Submitters have *higher* proportions of Need-Not-Met-With-Gift-Aid (and No-Need, No-Aid), and *lower* proportions of No-Need students who received "merit" aid. The gap closes slightly between the admit and enrolled stages of the funnel.

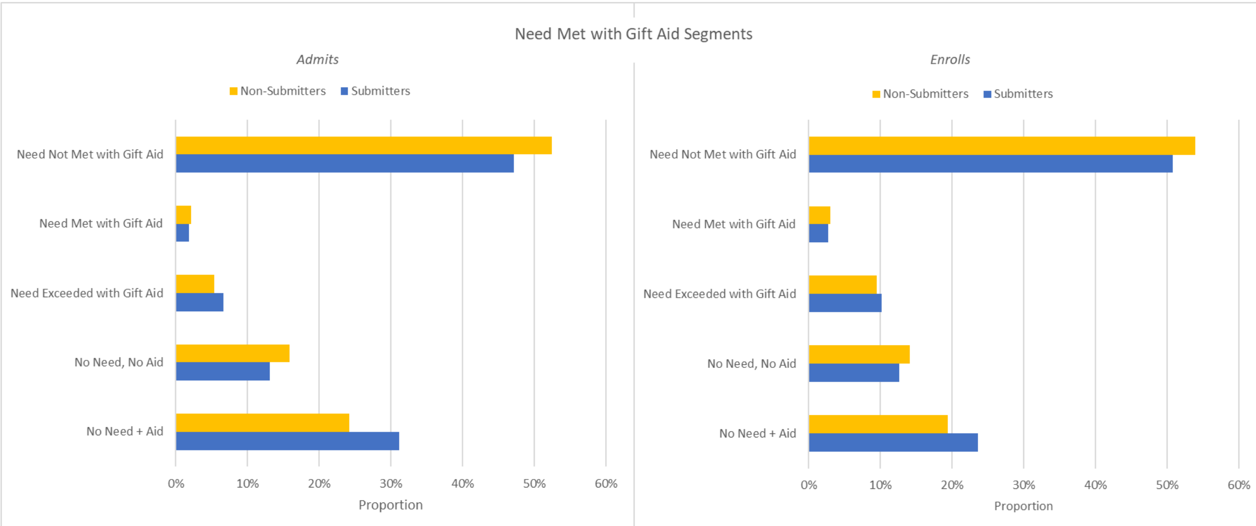


Figure 42. Need Met Segment Proportions by Phase of the Funnel and Submitter Status. Pooled Student Data. Subset of institutions selected based on availability of data at both the admit and enroll level, offering merit aid programs.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens. No public institutions provided data at the admit level.
Admits: 10 Institutions **|** 77,888 Records (59,126 Submitters and 18,762 Non-Submitters)
Enrolls: 10 Institutions **|** 14,681 Records (10,578 Submitters and 4,103 Non-Submitters)
Enrollment = 1,500-7,000 (2,400 M) **|** Endowment per FTE = $40,000-$500,000 ($63,000 M) **|** Admit Rate = 30%-70% (50% M)

These charts were limited to the ten institutions that were able to provide the data at both the admit and enrolled stage. The next chart, illustrates the same comparison solely at the enrollee stage, but includes 15 institutions. Note that the pattern is virtually identical with 5 more institutions.

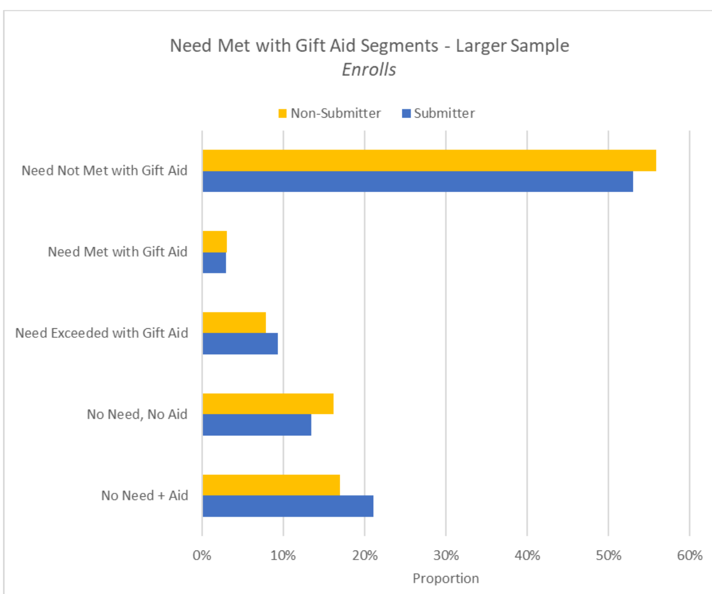


Figure 43. Need Met with Gift Aid, Enrolled Student Proportions by Submitter Status. Pooled Student Data. Private and public institutions selected based on availability of data the enroll level, offering merit aid programs.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
15 institutions **|** 25,798 Records
Enrollment = 1,500-20,000 (2,400 M) **|** Endowment per FTE = $6,000-$800,000 ($52,000 M) **|** Admit Rate = 15%-90% (43% M)

The next pair of charts (Fig 44) contrasts these same distributions for the enrolled students at a set of highly selective institutions in comparison with some less selective institutions. Note, at the less selective institutions, the larger proportion of Non-Submitters whose need was not fully met with Gift Aid, and on the other end of the spectrum, the larger proportion of No-Need Submitters that received Gift Aid. Also note on these charts that truly "full pay" students (No-Need, No Aid) that are less than 10% of the enrollees at the less selective institutions, comprise roughly 30% of the enrollees at the highly selective institutions. We do not know the policies that generate these results, but there are at least two reasonable hypotheses. First, the less selective institutions are not competitive enough to enroll many truly full-pay students, so they utilize "merit" scholarships more extensively (for both Submitters and Non-Submitters) to attract those affluent students. And second, the less selective institutions may have treated Submitters more generously Non-Submitters in the No-Need Plus Gift Aid, Need Exceeded, and Need Not Met With Gift Aid categories because they were pursuing profile-enhancing test scores in each of those cohorts, which obviously were not available for the Non-Submitters.

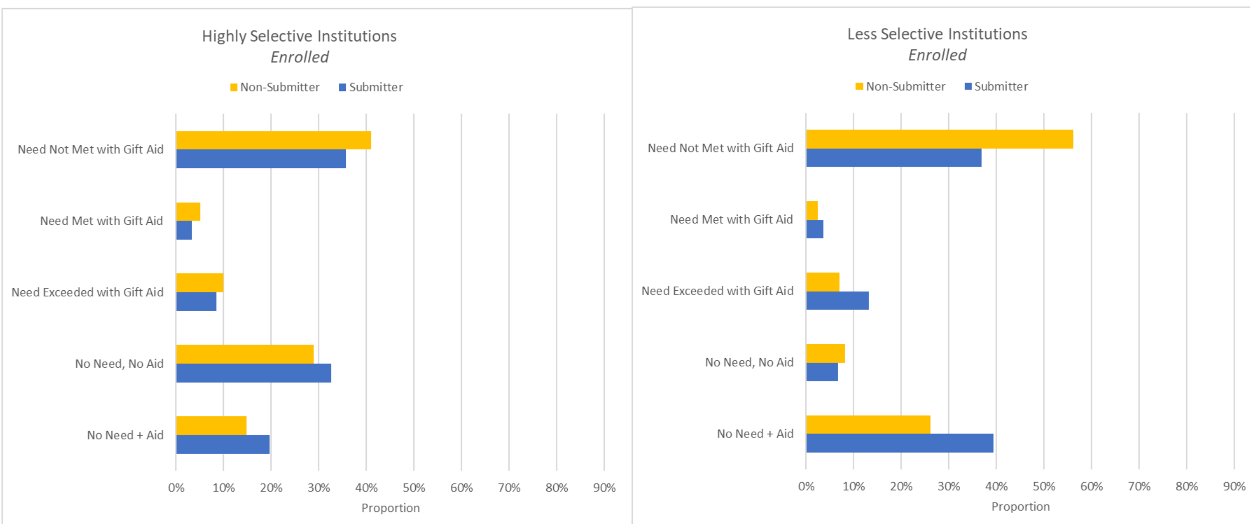


Figure 44. Need Met with Gift Aid, Enrolled Student Proportions by Submitter Status: Highly Selectivity vs. Less Selective. Pooled Student Data.
Highly Selective: Institutions with < 40% Admit Rate and reliable EFC and Gift Aid data at the Enroll level, offering merit aid programs.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
5 Institutions **|** 8,313 Records
Enrollment = 2,000-5,000 (3,000 M) **|** Endowment per FTE = $10,000-$230,000 ($180,000 M) **|** Admit Rate = 20%-40% (34% M)
Less Selective: Institutions with > 60% Admit Rate and reliable EFC and Gift Aid data at the Enroll level, offering merit aid programs.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
5 Institutions **|** 13,445 Records
Enrollment = 1,500-10,000 (4,000 M) **|** Endowment per FTE = $20,000-$60,000 ($38,000 M) **|** Admit Rate = 60%-70% (64% M)

An interesting side note is that very few of our participating institutions currently require test scores of merit scholarship candidates, whereas a number of them did in earlier years (perhaps including the cohorts for which there is data in this study). As institutions become more comfortable with their use of TOP, it may be that they feel less compelled to require test scores for merit scholarship consideration.

Are NEEDY Non-Submitters treated differently than Submitters in gift aid allocation?

There were substantial differences in awarding strategies between institutions, with the majority having provided less generous (gift) aid packages to their <u>needy</u> Non-Submitters than to their <u>needy</u> Submitters.

The chart below provides comparisons of the financial aid treatment of Submitters versus Non-Submitters. All the data are averages, so they do not account for differences that may have occurred in the awarding of high-need versus low-need students, but they do suggest that only a couple of institutions appear (on average) to have been more generous with Non-Submitters than with Submitters, while the majority of these institutions appear to have treated Submitters more favorably.

	Expected Family Contribution (EFC) *Federal or Institutional Definition*		Family Financial Contribution (FFC) *Total Cost of Attendance - Gift Aid*		DELTA			SUMMARY
Institution	**Non-Submitter**	**Submitter**	**Non-Submitter**	**Submitter**	**Non-Submitter Delta** *EFC-FFC*	**Submitter Delta** *EFC-FFC*	**DIFF** *NS-S*	
Private, Less Selective	$ 9,653	$ 13,652	$ 26,680	$ 23,392	$ (17,027)	$ (9,740)	$ (7,287)	>>Non-Submitters treated less favorably than Submitters
Private, Less Selective	$ 7,511	$ 13,217	$ 15,155	$ 15,852	$ (7,644)	$ (2,636)	$ (5,008)	
Private, Less Selective	$ 19,111	$ 23,737	$ 37,624	$ 37,943	$ (18,513)	$ (14,206)	$ (4,307)	
Private, More Selective	$ 16,123	$ 22,670	$ 19,089	$ 22,639	$ (2,966)	$ 31	$ (2,997)	
Private, More Selectivity	$ 19,087	$ 24,198	$ 27,810	$ 30,011	$ (8,724)	$ (5,814)	$ (2,910)	
Private, Moderately Selective	$ 15,100	$ 18,064	$ 22,463	$ 22,839	$ (7,363)	$ (4,775)	$ (2,588)	
Public, Less Selective	$ 6,825	$ 9,603	$ 21,060	$ 21,447	$ (14,235)	$ (11,844)	$ (2,391)	
Private, Less Selective	$ 7,391	$ 12,123	$ 21,238	$ 23,745	$ (13,847)	$ (11,622)	$ (2,226)	
Public, Less Selective	$ 4,323	$ 5,691	$ 16,684	$ 15,834	$ (12,361)	$ (10,143)	$ (2,219)	
Private, Less Selective	$ 17,005	$ 22,671	$ 28,946	$ 32,492	$ (11,941)	$ (9,821)	$ (2,120)	
Private, More Selective	$ 18,565	$ 19,017	$ 36,657	$ 35,580	$ (18,092)	$ (16,564)	$ (1,528)	
Private, More Selective	$ 18,544	$ 22,837	$ 27,096	$ 30,183	$ (8,552)	$ (7,346)	$ (1,206)	
Private, Moderate Selectivity	$ 17,369	$ 19,388	$ 25,447	$ 26,287	$ (8,077)	$ (6,900)	$ (1,177)	
Private, Less Selective	$ 22,487	$ 24,299	$ 39,303	$ 40,158	$ (16,816)	$ (15,859)	$ (957)	>>Equivalent treatment of Submitters and Non-Submitters
Private, Moderate Selectivity	$ 12,975	$ 18,303	$ 23,092	$ 27,966	$ (10,117)	$ (9,663)	$ (454)	
Private, More Selective	$ 15,422	$ 19,381	$ 21,313	$ 25,799	$ (5,891)	$ (6,418)	$ 527	
Private, More Selective	$ 18,719	$ 20,851	$ 19,945	$ 22,939	$ (1,225)	$ (2,088)	$ 862	
Private, More Selective	$ 21,236	$ 23,998	$ 25,607	$ 30,318	$ (4,372)	$ (6,321)	$ 1,949	>>Non-Submitters treated more favorably
Private, More Selective	$ 15,168	$ 20,574	$ 21,049	$ 28,470	$ (5,881)	$ (7,895)	$ 2,014	

Color Key	
No Difference	< 0.1
Small Difference	0.1 - 0.3
Medium Difference	0.3 - 0.5
Large Difference	> 0.5

Figure 45. Average Demonstrated Need Differential: Non-Submitter vs. Submitter. Private and public institutions with reliable EFC and Gift Aid data on enrolled students with a qualified need.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
19 institutions **|** 19,307 Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$230,000 ($60,000 M) **|** Admit Rate = 20%-80% (50% M)

Are NO-NEED Non-Submitters treated differently than Submitters in gift aid allocation?

In short, yes. Potentially exacerbated by the policy at some institutions to require test scores for scholarship consideration, we found that no-need Non-Submitters were awarded gift aid at lower rates than no-need Submitters. And yet these Non-Submitters graduated at modestly higher rates than their Submitter counterparts.

Considering only the Admitted, No-Need students, the next table compares the proportions of the Non-Submitters and Submitters that received gift aid. In this chart, as in others, we use the term "merit" broadly to refer to gift aid that does not appear to have been awarded based upon financial need. As noted elsewhere, a small percentage of these students may actually have received (for instance) non-institutional grants (e.g., Pell) based upon the federal methodology, but have been deemed "no-need" based upon institutional methodology. Our data did not allow us to distinguish these differences. In the pooled data from these 10 institutions, No-Need Submitters were awarded gift aid at a higher rate (70%) than their Non-Submitter counterparts (60%). As noted above, for the cohorts included in this study, some of these institutions may have required test scores of anyone seeking consideration for a merit scholarship, and if so, it would have contributed to this disparity.

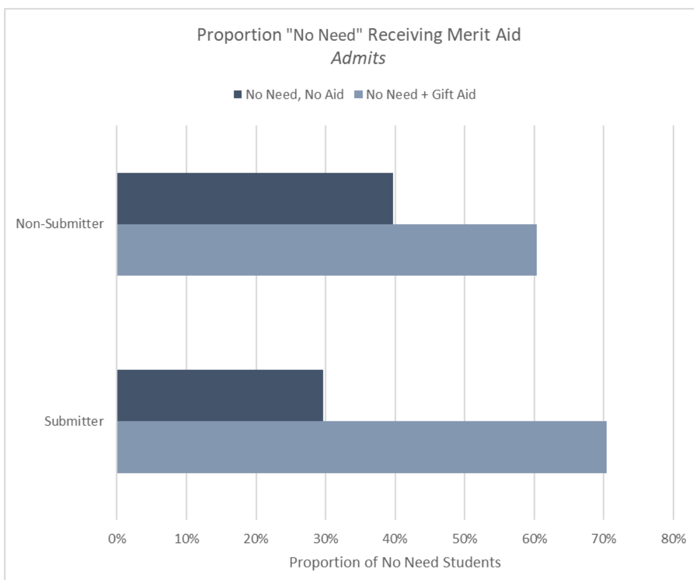


Figure 46. Focus on Admitted "No Need" Segments. Pooled Student Data. Institutions with admit level data, merit aid programs, reliable financial data. Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens. No public institutions provided data at the admit level.
10 Institutions **|** 33,718 Records (students with "No Need")
Enrollment = 1,500-7,000 (2,400 M) **|** Endowment per FTE = $40,000-$500,000 ($63,000 M) **|** Admit Rate = 30%-70% (50% M)

As with many of our comparisons, the pooled data is helpfully augmented by seeing the disaggregated data presented by individual institutions. For the same data as in Fig 46 above, the chart below illustrates the difference, at the institution level, between the proportions of No-Need Non-Submitters and Submitters who received non-need-based grants or scholarships, with each bar representing a separate institution. One institution appears to have treated No Need Non-Submitters more favorably, but the bulk of the institutions appear to have favored Submitters when granting grants and scholarships to Non-Need students. Virtually the same pattern is replicated within the enrolled students.

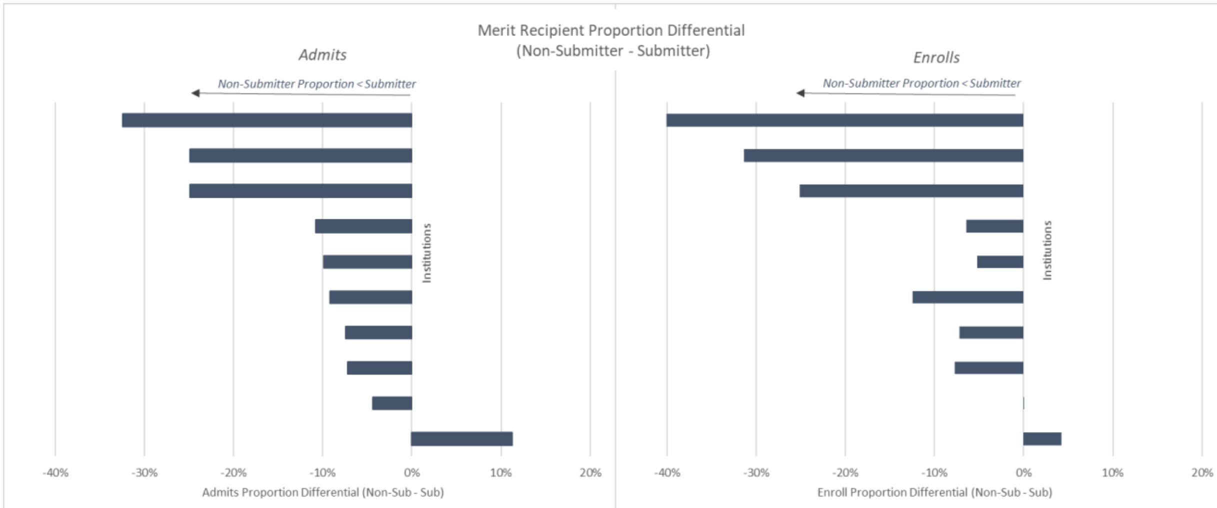


Figure 47. "No Need" or Merit Recipient Proportions Differentials: Non-Submitter vs. Submitter. Subset of institutions selected based on availability of data at the admit level and enroll levels, and institutions with merit aid programs. *Institutions are aligned in the two tables.* Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens. No public institutions provided data at the admit level.
10 Institutions**|** 33,718 Admit Records **|** 5,205 Enroll Records (students with "No Need")
Enrollment = 1,500-7,000 (2,400 M) **|** Endowment per FTE = $40,000-$500,000 ($63,000 M) **|** Admit Rate = 30%-70% (50% M)

As noted above, these no-need or "merit" awards may be given for a variety of reasons, from the desire to enhance the institutional profile by enrolling high test-score students (implicitly excluding Non-Submitters) to a tacit acknowledgement that the institution cannot enroll the right mix of needy and full-pay students at their published price, and therefore need to discount that price for some no-need students. The situation for the Non-Submitters is exacerbated at the institutions that specifically require the submission of test scores to receive consideration for their merit scholarship programs. But the effect is the same: the families of higher need students (including a higher proportion of Non-Submitters), are carrying more "self-help" and smaller grants relative to their need, while no-need students (including a higher proportion of Submitters) are given merit awards to reduce the price to the family. At some institutions, the net financial expectations for families with no need and high need were within a few thousand dollars of each other, because of the combinations of merit awards and high levels of self-help packaging. But, to summarize, with regard to aid awards at these institutions, on average, Non-Submitters didn't fare as well as Submitters.

How does aid allocation relate to graduation rates?

*With respect to graduation rates, we found that No-Need, Non-Submitters were **less likely** to receive gift awards, but they graduated at marginally higher rates than the No-Need Submitters.*

The chart below presents the graduation rates at the 14 institutions that have sufficient aid data and have been TOP long enough to have graduation rates. At least within this small sample, it appears that the lowest graduation rates were achieved by students whose need was not met fully with gift aid – not surprising, as they were likely the students facing the greatest financial challenges to completing their degrees. Non-Submitters whose need was fully met with Gift aid and those who were no-need and received no aid, graduated at modestly lower rates than their

Submitter counterparts. However, No-Need Non-Submitters who were awarded gift aid graduated at a modestly higher rate than No-Need Submitters. And yet the institutions favored the Submitters in granting this aid, which may be a counter-productive strategy.

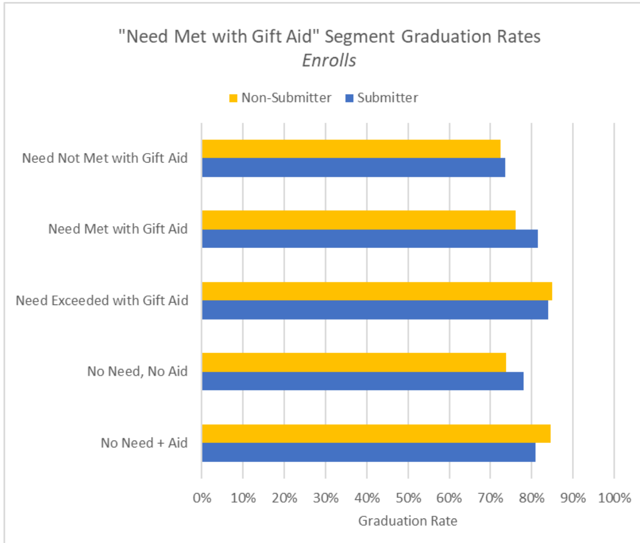


Figure 48. Need Met Segment Graduation Rates, Submitter vs. Non-Submitter. Pooled Student Data. Public and private of institutions selected based on availability of Gift Aid data at the enroll level, and institutions with 4-5 Year graduated TOP cohorts.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens.
14 Institutions **|** 17,436 Records
Enrollment = 1,500-5,000 (2,400 M) **|** Endowment per FTE = $6,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (40% M)

Figure 49 below provides another optic, examining the graduates and non-graduates of 17 institutions with 4 or 5 year graduation rates. The non-graduate Submitters and Non-Submitters appear to be similar in both their high school profiles and in their college academic records – modestly but not dramatically weaker than their graduating peers. On average it does not appear that they were required to leave for academic reasons. The non-graduates have higher representations of URM, First Gen, and Pell recipients. But we note a painful reality: comparing the four categories of graduates and non-grads, Non-Submitters and Submitters, the graduate Submitters have both the *lowest* demonstrated need and the *lowest* Family Financial Contribution. In contrast, the non-graduate Non-Submitters have both the *highest* demonstrated need and the *highest* Family Financial Contribution. And supporting the observation in the preceding chart, the Need-Not-Met students comprise a substantially larger share of the non-graduates than of the graduates. These higher financial expectations suggest that these students may be facing additional challenges to successfully navigating their college careers.

	Graduates All graduated students that entered under TOP policy cohort years 2011, 2012		Non-Graduates All non-graduated students that entered under TOP policy cohort years 2011, 2012. (This group of students has had 4-5 years to graduate, but have either withdrawn or are still enrolled.)	
	Non-Submitters	**Submitters**	**Non-Submitters**	**Submitters**
n	4311	12525	1377	4021
High School GPA	3.62	3.66	3.42	3.48
Academic Rating	6.05	6.69	5.65	6.04
SAT	1115	1249	1057	1177
First Year GPA	3.11	3.29	2.82	2.90
Cum GPA	3.25	3.40	2.85	2.89
URM	17%	10%	19%	13%
First Generation	15%	11%	20%	14%
Gender (Female)	63%	59%	61%	53%
STEM	23%	32%	12%	16%
Pell	21%	17%	31%	26%
Demonstrated Need	$40,026	$35,576	$41,692	$36,533
Family Financial Contribution	$37,528	$36,450	$39,002	$37,420
	In order to view proportions related to gift aid allocation, the rows below represent a subset of 11 institutions with graduating TOP cohorts and merit aid programs.			
n	2818	8400	992	2666
Exceed Need	10%	11%	6%	7%
No Need + Aid	17%	21%	19%	21%
No Need, No Aid	23%	20%	17%	14%
Need Not Met	46%	45%	55%	55%
Need Met	4%	3%	3%	2%

Academic Rating: All institutions submitted their respective scales, but for comparison purposes we converted all to a 10 point scale, where 10 is the highest rating.

SAT Submissions for Graduates: 10,747 *Submitters*, 1,243 *Non-Submitters*

SAT Submissions for Non-Graduates: 3,660 *Submitters*, 340 *Non-Submitters*

Figure 49. Academic Outcomes and Financial Contributors: Graduates vs. Non-Graduates. Enrolled students at **17** institutions with either 4 or 5 year graduation rates. Please note that the analysis is valid in comparing the Submitters versus the Non-Submitters for each line item, but each variable represents a different combination of institutions and students.
Exclusions: Two with Academic Threshold policies, one that did not have Non-Submitter proportions, and Non-Resident Aliens (excluded from the bottom section).
17 institutions **|** 22,234 Records
Enrollment = 1,500-5,000 (2,400 M) **|** Endowment per FTE = $6,000-$800,000 ($100,000 M) **|** Admit Rate = 15%-90% (40% M)

How does aid allocation relate to any gains in diversity?

With respect to diversity, to achieve the goal of serving more traditionally-underserved populations, we found that institutions will likely need to make additional financial aid commitments.

With the adoption of a TOP, the bulk of the schools in this study appear to have increased their enrollment of students from traditionally underrepresented (and generally needier) groups. We examined the financial impact that may have been required to secure these gains. In this section we have aligned financial aid averages against the proportions and growth of the three identifiable populations we have used earlier: First-Generation College Bound, students from lower SES backgrounds (using Pell Grant recipients as an easily identifiable proxy), and students from racial and ethnic groups that have traditionally been underrepresented in college populations (URM). Discussing them as three unique populations fails to account for the overlap or intersectionality of the three. Thus, as we did earlier, we have utilized the construct of "Expanded Diversity" in which we have included any student identified with any of these three groups.

In the study we collected only the total gift aid awarded to students and an indication of whether or not they had received a Pell Grant. Because it was too difficult for institutions to supply the specific breakdown of institutional gift aid, we are unable to draw solid conclusions about the financial impact of the policy on the institution. However, we devised an approximation by summing the Total Gift Aid for every enrolled student (regardless of whether it was need-based or merit) and then dividing that by the total number of enrolled students (whether or not they received aid) to arrive at the “Gift Aid Per Capita.” Obviously, some of this gift aid was not institutional gift aid (e.g., Pell Grants, State grants, and outside scholarships), but it does allow for a very rough assessment of the relative financial investment made to the pre- and post-policy-adoption cohorts.

We reviewed detailed information about the 12 institutions that were able to submit reliable financial aid data immediately pre-and post-policy-adoption in Figure 50 below.

			Private 1	Private 2	Private 3	Private 4	Private 5	Private 6	Private 7	Public 1	Public 2	Private 8	Private 9	Private 10
FINANCIAL AID DATA	**Qualified Need Proportions** *Enrolled*	PRE	44%	34%	41%	74%	44%	52%	69%	81%	75%	77%	64%	76%
		POST	56%	37%	51%	74%	48%	56%	67%	81%	75%	67%	61%	68%
		DIFF	12%	3%	9%	0%	5%	4%	-2%	0%	-1%	-10%	-4%	-8%
	Average Demonstrated Need --Needy Students-- *Enrolled*	PRE	$ 29,071	$ 34,177	$ 37,106	$ 39,464	$ 45,431	$ 41,853	$ 35,598	$ 24,180	$ 21,389	$ 35,372	$ 28,716	$ 42,942
		POST	$ 34,812	$ 38,786	$ 39,503	$ 45,602	$ 48,784	$ 41,582	$ 37,974	$ 25,582	$ 22,731	$ 34,836	$ 30,414	$ 40,308
		DIFF	$ 5,740	$ 4,609	$ 2,397	$ 6,138	$ 3,352	$ (272)	$ 2,376	$ 1,401	$ 1,343	$ (536)	$ 1,698	$ (2,634)
	Gift Aid *Per Capita* --All Students-- *Enrolled*	PRE	*No Data Submitted to Study*	$ 11,306	$ 14,082	$ 25,127	$ 17,367	*No Data Submitted to Study*	$ 19,102	$ 10,704	$ 5,106	$ 20,850	$ 10,477	$ 15,436
		POST		$ 13,505	$ 17,884	$ 27,461	$ 19,620		$ 22,539	$ 11,432	$ 6,894	$ 22,819	$ 10,407	$ 17,238
		DIFF		$ 2,199	$ 3,802	$ 2,334	$ 2,253		$ 3,437	$ 728	$ 1,788	$ 1,969	$ (71)	$ 1,803
	Average Gift Aid -- Needy Students -- *Enrolled*	PRE	*No Data Submitted to Study*	$ 27,526	$ 33,451	$ 28,171	$ 41,654	*No Data Submitted to Study*	$ 24,601	$ 12,544	$ 6,953	$ 25,197	$ 13,308	$ 26,753
		POST		$ 32,316	$ 36,297	$ 30,734	$ 44,486		$ 28,027	$ 13,901	$ 7,769	$ 29,288	$ 13,599	$ 24,752
		DIFF		$ 4,790	$ 2,846	$ 2,563	$ 2,832		$ 3,426	$ 1,357	$ 815	$ 4,091	$ 291	$ (2,001)
DIVERSITY OUTCOMES	**URM** *Enrolled Proportions*	PRE	10%	11%	9%	13%	16%	10%	12%	16%	6%	10%	8%	13%
		POST	19%	13%	14%	16%	18%	11%	15%	19%	4%	13%	6%	9%
		DIFF	*9%*	*2%*	*5%*	*3%*	*2%*	*1%*	*3%*	*3%*	*-1%*	*3%*	*-3%*	*-4%*
		% Change	*85%*	*17%*	*48%*	*23%*	*12%*	*13%*	*29%*	*19%*	*-22%*	*36%*	*-34%*	*-29%*
	Pell *Enrolled Proportions*	PRE	11%	7%	9%	27%	16%	15%	21%	28%	30%	24%	18%	14%
		POST	19%	10%	15%	32%	18%	17%	20%	28%	31%	22%	13%	13%
		DIFF	*7%*	*3%*	*5%*	*5%*	*2%*	*1%*	*-1%*	*0%*	*1%*	*-2%*	*-4%*	*0%*
		% Change	*65%*	*42%*	*56%*	*18%*	*11%*	*9%*	*-3%*	*0%*	*2%*	*-8%*	*-25%*	*-1%*
	First-Generation-to-College *Enrolled Proportions*	PRE	5%	5%	12%	24%	14%	13%	12%	34%	22%	23%	*No Data Submitted to Study*	21%
		POST	10%	7%	14%	31%	16%	15%	11%	36%	21%	21%		17%
		DIFF	*6%*	*2%*	*2%*	*8%*	*2%*	*2%*	*-1%*	*2%*	*-1%*	*-1%*		*-3%*
		% Change	*117%*	*31%*	*21%*	*33%*	*11%*	*16%*	*-8%*	*5%*	*-3%*	*-6%*		*-17%*
	Diversity, Expanded *Enrolled URM, Pell and or FG-- each student counted once*	PRE	21%	19%	18%	43%	29%	29%	31%	51%	43%	39%	*22%*	32%
		POST	35%	22%	25%	48%	31%	30%	32%	52%	44%	38%	*15%*	29%
		DIFF	*14%*	*3%*	*7%*	*5%*	*2%*	*1%*	*0%*	*1%*	*0%*	*-1%*	*-7%*	*-2%*
		% Change	*67%*	*19%*	*38%*	*12%*	*6%*	*4%*	*1%*	*2%*	*1%*	*-1%*	*-30%*	*-7%*

Figure 50. Table of Financial Aid Variables Pre-Post Policy, for Enrolled Students. Data from 12 private and public institutions with reliable financial aid data at the enroll level, submitted immediately pre- and immediately post-policy adoption.
Exclusions: Early adopters and Non-Resident Aliens
12 institutions **|** 38,047 Records
Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$230,000 ($65,000 M) **|** Admit Rate = 20%-80% (53% M)

We offer three example case studies of the interplay between financial investment and growth in diversity.

<u>Case Study 1</u> - This private institution achieved a 48% growth in the proportion of enrolled URM students (going from 9% to 14%, an increase of 5 percentage points), and also had one of the highest percentage increases in the proportion of needy students (going from 43% to 51%). The average need rose $1,868 and it appears they may have become more generous in their aid award policy, as the Gift Aid Per Capita increased by $3,802. They also had substantive increases in Pell recipients and First-Generation students, resulting in one of the largest increases on our “Expanded Diversity” measure (increase of 7 percentage points – which for them represented a 38% increase in the proportion of these populations). So, they appear to have made a substantive financial commitment, in addition to adopting a TOP, and have achieved a substantial increase in the diversity of their class.

<u>Case Study 2</u> – On the other end of the spectrum is a private institution that posted a decline in the proportion of needy students and actually lowered its Gift Aid Per Capita a slight amount. But in doing so it suffered substantive declines in the proportions of both URM and Pell recipients.

Case Study 3 – More in the “middle of the pack,” is another private, with roughly the same proportion of needy students pre- and post-adoption. The average need of their enrolled students increased by about $3,900 and their Gift Aid Per Capita increased by about $2,200, but they also increased the proportion of their URM students by 3 percentage points, their First Gen students by 8 percentage points and their Pell Recipients by 5 percentage points resulting in a proportional increase of 12% (5 percentage points) in their Expanded Diversity score. So, subsequent to adopting a TOP, they invested more resources in aid and substantially increased the diversity of their student body, particularly in terms of serving First-Generation students.

“Apps increased but we didn’t back it up with a strong enough financial aid program. We recently strengthened our commitment to aid so it’s working better now.”

Dean from small private college

The chart below provides a visual summary of the conceptual ROI (Return On Investment). The higher an institution is on the vertical axis, the more proportional gain it has made in the diversity of its freshman class (based on our “Expanded Diversity” rating). And the further to the right it is on the horizontal axis, the more it has increased its investment in financial aid. (N.B. The dollars expressed on the horizontal axis should be interpreted as providing relative order of magnitude rather than precise numbers, as we do not have specific institutional gift aid available.)

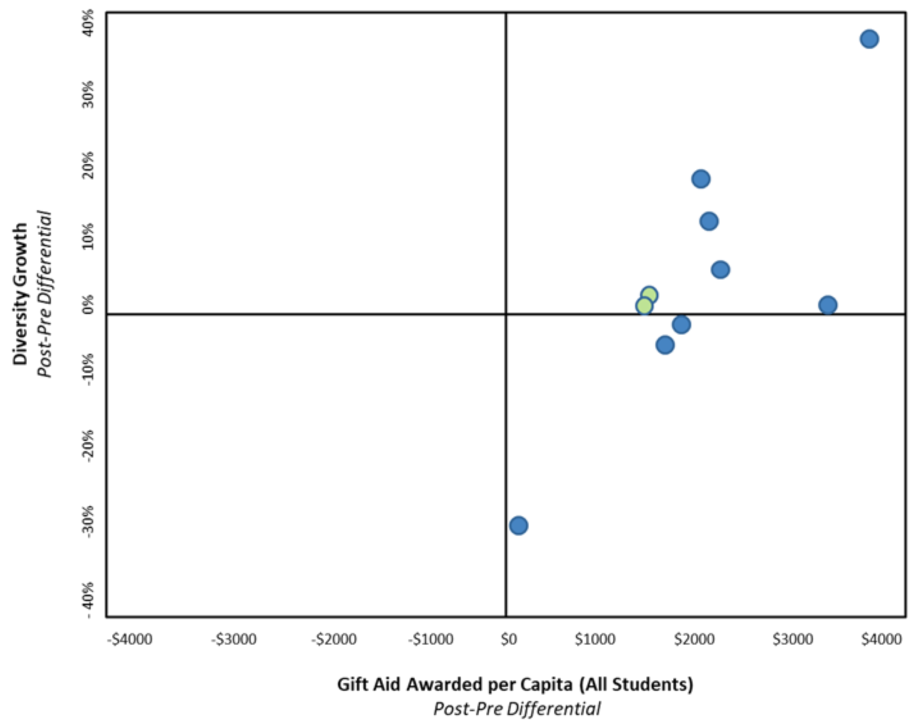


Figure 51. Investment-Outcome Matrix, Pre-Post Differentials of Expanded Diversity Percent Change and Gift Aid per Capita (All Students). Data from 10 private and public institutions with reliable gift aid data submitted immediately pre- and immediately post-policy adoption. Exclusions: Early adopters and Non-Resident Aliens

Points in green are public institutions.

10 institutions **|** 41,429 Records

Enrollment = 2,000-20,000 (4,000 M) **|** Endowment per FTE = $4,000-$230,000 ($65,000 M) **|** Admit Rate = 20%-80% (53% M)

So, while it is not possible (based upon this small sample of ten institutions) to arrive at definitive conclusions, it is reasonable to state that if an institution is interested in increasing its diversity, a test-optional admission policy can provide a valuable tool, but, unsurprisingly, to achieve the goal of serving more traditionally-underserved populations, the institution will likely need to make additional financial aid commitments.

SUMMARY AND IMPLICATIONS

What can we summarize about the workings of a Test-Optional Policy?

No research can provide definitive answers to the questions we have explored about the use of a test-optional admission policy, and we do not claim to have done so. Every institution has a slightly different experience. We are not presumptuous enough to believe we can identify any single outcome (nor even strong tendency) that should be expected by <u>every</u> institution adopting a TOP. There is simply too much variation between institutions in commitment to mission, location in the educational marketplace, student population being served, affluence, and institutional goals. Instead, we have provided as many perspectives on the <u>possibilities</u> that a TOP may help an individual institution to achieve.

Thus qualified, our findings suggest that a TOP works well at a variety of institutions. Almost all institutions in our study increased enrollment of underserved populations, with many showing proportionate increases exceeding those found at test-requiring peer institutions. And, the policy transition occurred without any signs of academic slide: GPAs and graduation rates didn't suffer, and according to reports from the Deans many faculty were very pleased with the quality and character of the incoming classes.

This success, however, appears to come with some degree of additional financial investment. The proportion of needy students rose at roughly half of our TOP institutions. Almost all the institutions saw an increase in the average demonstrated need from the pre-policy to the post-policy cohort years and increased the gift aid per capita. Most of our participating institutions appear to have provided less generous gift aid packages to Non-Submitters (both needy and no-need) than to their Submitters.

The institutions in our study appeared to treat Non-Submitters differently than Submitters, admitting them at a lower rate and, on average, treating them a little less generously in the financial aid process, particularly with merit scholarships. The admitted Non-Submitters, however, enrolled at higher rates at virtually all of our institutions.

These Non-Submitters comprised significantly larger proportions of Underrepresented Minorities, First-Generation-to-College, Pell recipients, and women than did Submitters. As a group, Non-Submitters showed slightly lower academic performance both in high school and college, but graduated from college at equivalent, or in some cases, higher rates than Submitters. The largest distinguishing academic difference we found was the lower test scores for the Non-Submitters (though we had test scores for only about a third of them).

Furthermore, this study helps to punctuate the question of what is meant when we refer to "success in college," a phrase frequently used to argue for the predictive value of the SAT and ACT. There is general agreement that those tests, when used in conjunction with high school grades, do a marginally better job than high school grades alone of predicting the First Year College GPA of students. However, whether they predict evenly across populations of students has been widely debated. And an increasing number of voices are challenging the notion that predicting whether a student is likely to achieve, say a 3.3 GPA versus a 3.2 at the end of their first year in college is synonymous with predicting "success in college," and are rejecting that phrase as an obfuscation of the actually limited value of the tests. Quoting again from the NACAC Commission on the Use of Standardized Tests in College Admission:

Commission members unanimously agreed that college success is a term of sufficient breadth that it includes degree attainment, a wide range of GPAs, and the acquisition of experiences and skills that will propel a student into the workforce, graduate education, or responsible citizenship. (NACAC, 2008)

We also found that this group of Non-Submitters represented approximately a quarter of the applicant pool, a finding that highlights an interesting intersection between findings published by the testing agencies. Their meta-analysis of studies also found a quarter to a third of all students with "discrepant performance," either students with high HSGPAs and low testing, or the reverse: modest HSGPAs but high testing (Sanchez & Mattern [Making the Case for Standardized Testing] in Buckley, 2018). Within this group, approximately half of them have high school grades that are higher than what the tests would predict. And it is worth noting the parallels between the proportions noted in the above studies and the median percentage (23%) of students choosing to be Non-Submitters at the institutions in this study.

We also agree with characterizations of the most likely students to have strong HSGPAs and low testing: women, First-Generation to college, low income students, and students who speak a second language at home.[21] Furthermore, those most likely to be discordant with weaker HSGPAs and stronger testing are males, whites, and those of higher income (Sanchez and Edmunds, 2015).

We would suggest another, largely parallel language for thinking of these students. Many researchers, especially in medical fields, will speak of "false negatives" and "false positives." A false negative occurs when the test suggests that something will not happen, but it does. A false positive suggests that something will happen, but it doesn't. We assert that most TOP Non-Submitters are "false negatives": the SAT and ACT tests suggest that they will not perform well in college, but these students perform fine, and graduate at equal or higher rates than the Submitters.

Finally, this study also confirmed that the SAT and ACT do have a positive correlation with college cumulative GPA for some students, more commonly Submitters -- the students who made an informed decision that their testing represented their ability. We do not argue that institutions should entirely eliminate consideration of the ACT and SAT for all their students. We do not promote the simplistic notion that these tests are either "all bad" or "all good."

The argument from the testing agencies that colleges should want every piece of significant information to make their decisions misses a key point. A student's decision to apply to TOP colleges, and not to have test scores considered in the admissions decisions, is significant information, often profoundly important for both student and institution. The students have made a key decision, saying to the Admissions offices, "I am a better student and potential citizen than these tests would suggest." The research findings from this study and others cited suggest that the students are right.

[21] Hiss and Doria, 2010. A 25-year study of optional testing at Bates found a thick band of Non-Submitters whose homes ran across the top of ME, NH and VT. Often American citizens for several generations, they were of French Canadian backgrounds. Being close to the border, they had kept up cultural and linguistic ties, with students speaking French at home and learning English at school.

What did TOP deans say about their experience with the policy?

At the start of the study, we interviewed each participating Dean of Admission about their experience with TOP. Then, we reviewed their comments in light of the data submission, which was subject to a thorough analysis. Pulling apart some of these unique findings from each institution enabled us to see patterns in similar experiences.

- The motivations cited for adopting a test-optional policy were fairly consistent across the institutions in the study, relating primarily to **improving the access of underserved students**:
  - *"TOP arose from the decision to pursue access to higher education among underrepresented groups."*
  - *"It was important to our commitment to access, in particular to First-Generation students and students from under-resourced schools. Also felt it might help us reach students who might previously not have considered."*
- The adoption of the policy was also described as way to simply **formalize what they had been doing in practice all along**. *"We never weighted testing heavily – always weighted classroom performance more heavily"* or another institution *"We always pitched that the scores were not given much weight."*
- While the policies varied in terms of specifics, most employed an **Optional for All** policy with a few exceptions, for example, requiring test scores from international students, homeschooled students, or students applying to specialty programs. Some had started with an Academic Threshold or Optional Plus policy, and then migrated to the more open variation of the policy, Optional for All, indicating an increased comfort level with making sound admission decisions without testing.
- The group that employed some form of an **Optional Plus policy had mixed reviews**. A few seemed pleased with their interviews, essay questions or other formats designed to solicit information from students on non-cognitive skills. However, others felt that it was too time consuming and did not yield the results they wanted, *"The [additional requirement] added very little to our reading or prediction, and took up big shares of time."* These institutions are actively considering eliminating the additional admission materials and migrating to an Optional for All policy.
- Most of the institutions **had not employed a marketing campaign** to promote their new policy. Many claimed to simply *"Put out a press release and some FAQs on website"* or *"launched it by notifying on the Common App and on the website, and included it in a counselor update newsletter."*
- A few institutions **launched TOP as part of a larger initiative** "*TOP was a component of a substantive shift of the institution as more hands-on skill building, and felt standardized tests did not fit well with this shift."*
- Most described the policy buy-in process as taking some time and effort, particularly with faculty members. Yet, many claimed that after TOP adoption **constituents reported being satisfied with it**. As one dean claimed *"We are very happy with the policy and there is no discussion of rolling it back. We also listen heavily to the comments of faculty about their students and hear no concerns."*
- Many described the process of transitioning to TOP as largely painless, just requiring an additional step in the direction of **focusing greater attention on the rigor of the coursework in the context of the quality of the high school**. One noted, *"a little more training with our readers, and it has increased our focus on the rigor of the high school curriculum."*
- Some institutions mentioned adding interviewers or temporary readers to handle the application increase. One of the institutions, as highlighted in the introductory case study, hired **additional temporary readers** because of the increased volume of applications, and the additional time required for Non-Submitter applications. *"When you have a test score that is consistent with the academic record and other documents, it gives you confidence to spend less time on the application. When there's no test score as 'confirming' evidence, you'll typically look more closely at the high school record, the rigor of the curriculum, and the school profile for context."*

- Most of the Deans described their **policy experience in glowing terms** as a success in achieving admission goals:
    - *"We are attracting more Students-of-Color due to the policy."*
    - *"We are seeing lots of kids who have done everything right except have high tests."*
    - *"It has absolutely worked. First year academic performance and freshman-sophomore retention have improved. We've seen a steady increase in ethnic diversity."*
- A few had more limited success, and had to make some adjustments to the policy over time:
    - *"Only issue that did not play out as well as we had hoped with students of color, First Gen, and working-class kids. But we have fixed the financial aid budget now."*
    - *"It has worked, though it is not nearly as popular (widely used) as we thought it would be…maybe most of the students who would traditionally have been attracted to TOP were already applying without being overly concerned if their test scores didn't represent them well."*

What are the authors' ideas for future research?

This study has provided many insights into the use of a test-optional policy. However, while our study is both broad and detailed, we identified numerous related topics that reasonably could be addressed in future research, including in no particular order of priority or importance:

- Our study focused on analyzing data from student records provided by institutions, and did not delve into the nature and behavior of Non-Submitters. Qualitative research techniques, such as interviews and focus groups with prospective students and enrolled Non-Submitters, would be an insightful follow-up to this study.
- While graduation rates were a pivotal feature of this study, the specifics of student retention were not investigated. Much could also be learned about the success of the policy in light of the on-campus learning and social environment that receives Non-Submitters.
- We received test scores for a limited proportion of Non-Submitters. A more comprehensive collection of test scores from Non-Submitters would allow a more definitive assessment on the alignment of HSGPA and test scores with college outcomes.
- LD student access to higher education is a future research project with potentially very high rewards, given the growing understanding and identification of this population of students.
- While we elicited complex findings on financial aid, there is much room for additional investigation, including the review of the policy in light of institutional aid numbers.
- Academic Threshold policies offer a degree of lessened reliance on testing, but are a very different breed of policy when compared against the Optional for All model, whereby students have the free will to choose their admission credentials. There is much room to learn more about the ins and outs of this type of policy.
- The concept we introduced of "Expanded Diversity," whereby each student is considered in terms of all the rich diversity he or she brings to campus life and learning, could be used to underpin a substantial rethinking about how we as researchers understand diversity in higher education. This concept has potential to encompass an even broader range of students than we did in this study, for instance, those with learning differences, non-traditional students, international students, and non-native speakers.

APPENDIX

CO-AUTHORS

Steve Syverson, Co-Author and Principal Investigator

Steve Syverson graduated from Pomona College, and worked in Admissions at Pomona (1978-1983), culminating as the Director. For 28 years until 2011 he served as the Dean of Admissions and Financial Aid at Lawrence University, substantially increasing the proportion of students-of-color and international students. He has designed and programmed admissions and financial aid computer systems for Pomona, Claremont McKenna, and Lawrence. In 2016, after a short stint as a retiree, in southern California, he accepted an appointment as the Assistant Vice Chancellor for Enrollment Management at the University of Washington Bothell.

Active in professional organizations, Steve has variously served on the NACAC Board of Directors (2011-2014), as NACAC Vice President for Admission Practices (1988-91), and on the NACAC Commission on the Use of Standardized Testing in Undergraduate Admissions which issued its highly influential report in 2008. He chaired the NACAC Media, Marketing and Technology Committee, and served on the faculties of the NACAC Guiding the Way to Inclusion and NACAC Tools of the Trade professional training workshops. He also was the President of Wisconsin ACAC and served on the ACT Executive Council of Wisconsin.

A frequent author and speaker on admissions and financial aid topics, he has published articles on "Basics of Standardized Testing in the Admissions Process," "The Role of Standardized Tests in College Admissions—Test Optional Admissions," "Ethics and Sensitivity to Students."

Outside of admissions, he has a particular interest in Habitat for Humanity and has worked extensively with the American Institute of Certified Educational Planners advancing the Certified Educational Planner (CEP) credential for independent educational consultants and school-based college counselors. He also has been a long-time member of the advisory board of Cappex.com.

Valerie W. Franks, Co-Author and Study Manager

Valerie Franks brings over twenty years of experience in research. In 2007, she founded her own firm to provide consulting and analysis to educational institutions. She has spent the past eight years researching Test-Optional Policies, first as Lead Researcher and co-author of the first study on Test-Optional Policy "Defining Promise" and now serving the same role in "Defining Access."

Prior to that, she was a former Assistant Dean of Admissions at Bates, where she recruited students, read applications and enrolled students in the context of a Test-Optional policy. Alongside this role, she also acted as admissions analyst, working closely with the Director of Institutional Research and the Dean of Admissions to examine admission funnel patterns, as well as report data to IPEDS.

Valerie's roots in research started in the business world. She spent two years as project executive in Prague for an international research firm, surveying the Central European market to identify market opportunities for international companies, with primary responsibilities for data analysis and formation of market strategy. She then spent five years as research manager for a New York strategy consulting firm, designing and managing national qualitative and quantitative research studies on brand equity, customer preference, competitor positioning and messaging resonance, brand identity, and positioning and creative / tactical execution plans.

She earned her BA in Psychology from Bates, and speaks fluent French and proficient Czech.

William C. Hiss, Co-Author and Advisory Committee Chair

Bill Hiss served Bates College for 35 years as Dean of Admissions and Financial Aid, a vice president supervising alumni relations, career services and communications, a senior leadership gifts officer, and a Lecturer in Asian Studies. His "Literature through Cataclysm" course studied the modern fiction and film of five societies that endured a 20th century cataclysm: Russia, Japan, Vietnam, India/Pakistan and Somalia.

Bill took his B.A. in English from Bates, an M.T.S. in ethics and American church history at Harvard Divinity School, and an M.A and Ph.D. at Tufts in American literature, religion and intellectual history. He taught in JHS 120 in the South Bronx, at Tufts as a graduate student and at Hebron Academy, as well as Bates. He served on the Federal Advisory Committee on Student Financial Assistance, which advises the Congress and Secretary of Education on national financial aid policy. He established and led the advisory committee of deans at USNews on guidebook and ranking issues.

In 1984 the Bates faculty made standardized testing optional for admissions. For over 35 years, Bill has researched and written on optional standardized testing, including a 25-year look-back study of the Bates optional testing policy with co-author Kate Doria. In February of 2014, Bill and his co-author Valerie Franks published the first national, peer-reviewed study, "Defining Promise," of optional testing at 33 private and public institutions. This study found only trivial differences between Submitters and Non-Submitters of testing in both Cum GPA's and graduation rates.

A retiree beekeeper, soccer referee, crew on sail and power boats, and lay UCC pastor, he recently spent a month in Ho Chi Minh City helping to design the Admission and Financial Aid offices for Fulbright University Vietnam, Vietnam's first liberal arts institution, and perhaps Asia's first TOP institution.

Lidia Ortiz assisted the research team in the data preparation and analysis. A graduate of Smith College with a bachelor's degree in Economics and Psychology, Lidia has previously collaborated with the University of Illinois at Chicago and the University of Southern Georgia Psychology department working on health disparities research. In addition, she has worked with Smith's College of Office of Institutional Research to enable data-driven decision making.

Members of the Advisory Committee

David Hawkins is the Executive Director for Educational Content and Policy at the National Association of College Admissions Counseling, where he has served for 18 years. He received his MA in Government from William and Mary.

Brian Prescott is the Associate Vice President at the National Center for Higher Education Management Systems. He has previously served a term as an appointed member of the NACAC Board of Directors and, while at the Western Interstate Commission for Higher Education, authored or co-authored two editions of WICHE's widely used projections of high school graduates, *Knocking at The Door*. He received his undergraduate degree from William and Mary, his M.A. from the University of Iowa, and his Ph.D. from the University of Virginia.

Kevin Rask has been College Research Professor of Economics at Colorado College since 2011. Before that he was Professor of Economics at Wake Forest University and Colgate University for 20 years. He received his undergraduate degree from Haverford and his Ph.D. from Duke, both in Economics. He has frequently published on higher education issues, including research on optional testing at Wake Forest as an essay in *SAT Wars: The Case for Optional Testing in Admissions.*

Cate Rowan is the Executive Director of Institutional Research at Smith College, where she has served since 2005. Previously, she was the Director of Research at Mt. Holyoke College for 8 years. She received her undergraduate degree from Smith, and her MBA from the Isenberg School of Management at the University of Massachusetts, Amherst.

DATA FIELDS DEFINED

Racial and Ethnic Student Identification Data

We used IPEDS-defined categories of racial/ethnic identification. For research purposes, we requested that institutions submit data using the hierarchical method of identification that counts each student only once. Some of our data spans the US Census and 2010 IPEDS change in the way race/ethnicity was recorded. We had no way to "correct for" that change, so Hispanic numbers, in particular, may have been confounded in comparisons that span those years, but in most instances we believe that including them in the overall Underrepresented Minority (URM) count reduces or eliminates the impact of the coding change. After much consideration, we did not include the IPEDS "Two or more races" categorization in our overall Underrepresented Minority grouping.

Instead, we found it helpful to design some analyses which looked at the overlap of various racial and ethnic groups. A genuine step of "forward motion" in college and university admissions in recent years has been the increasingly common practice of including data on First-Generation-to-College and Pell Grant recipients in presenting a class profile. We have followed that practice and tried to show the overlap between the various racial, ethnic, educational attainment, and income level groups. But we also have experimented a bit, creating some information on what we've termed "Expanded Diversity."

High School GPA Data

HSGPA data presented one of several interesting challenges as we attempted to use consistent data across the institutions in our study. As with all our data, HSGPAs were reported to us as recorded by the institution's Admissions or Institutional Research offices. Some colleges simply record whatever GPA the high school has supplied, whether weighted or unweighted, and whether or not it is a traditional 4-point scale. Other institutions follow internal protocols for converting HSGPAs to a common 4.0 scale. Many high schools record GPAs that exceed 4.0 as part of a weighting schema for honors, IB or AP classes, whereas other high schools do not weight their grades. Some colleges simply truncate anything above a 4.0 to a 4.0 and account for the rigor of the curriculum elsewhere in their process. So there is great variability in the treatment of HSGPAs both among the high schools and among the colleges, prompting us to wonder how other studies have attempted to normalize the treatment of HSGPAs. Studies that are internal to single institutions (e.g., regression analyses of the predictive value of grades or test scores) eliminate one layer of variability, but multi-college studies are particularly challenged.

We did an extensive individual analysis of the GPAs reported by each of the 28 institutions in the study, using whatever protocol that institution used, and allowing for the GPA scales to exceed the traditional 4-point scale. We also created a second HSGPA data element in which we truncated the GPA at 4.0 for each record. In the end, because many colleges and high schools did not report GPAs higher than 4.0, and there was no way to "unbundle" what was originally reported to the institutions by the high schools, we have used the truncated 4.0 methodology for our comparisons.

Standardized Test Score Data

We requested test scores on all student records. We received at most one set of ACT and/or SAT scores for any individual student. For simplicity of comparison and discussion, we used a concordance table[22] to convert ACT scores to SAT scores, and all references in the report refer to them simply as SATs. The new SAT had not been taken by most of the student cohorts in this study.[23]

We found that institutions either still received (or requested upon admission) test scores from students who chose to omit testing from their application, but at much lower rates than the 2014 study. Four institutions did not have scores on record for their Non-Submitters, and among the remaining 24 institutions, 27% of Non-Submitter applicant records contained a test score. At the enrolled student level, 36% of the non-submitter records had test scores, with a range of enrolled student Non-Submitter test submission across institutions of 7% to 82%. We represent and use this subset of scores with caution, as it does not represent the full range of scores from this student segment.

To test our findings at institutions with a higher rate of data collection, we conducted parallel analyses at 8 of our institutions that had SAT/ACT scores for at least 50% of their enrolling Non-Submitters, and found the same results as when we used all the institutions. It would seem to defy common sense to think that the scores which were not submitted were dramatically *higher* than those which the institutions did collect. If anything, one might assume that that the scores that were not collected, on average, would be lower. Yet with or without a collected test score, the Cum GPAs and graduation rates of the Non-Submitters speak to their ability to succeed in college and university curricula.

Financial Aid Data

We collected four principal data elements from the participating institutions:

- <u>Expected Family Contribution (EFC)</u> – At some institutions this was the federal EFC, and at others it was an institutionally-determined EFC. Because we were attempting to assess aid award differences between Submitters and Non-Submitters at the individual institution, we sought whatever EFC they used to determine the need for aid.
- <u>Total Gift Aid</u> – We asked only for the total gift aid (from any source) received by the student.
- <u>Pell Grant Recipients</u> – We asked institutions simply to identify any student receiving a Pell Grant, but did not ask for the specific amount of the Pell Grant.
- <u>Merit Aid Recipients</u> – We asked them to identify any student to whom they had awarded non-need-based, “merit” gift aid.

All financial aid values and institutional Costs-of-Attendance were adjusted into 2016 dollar equivalents[24] to be able to compare the values over time. We created a series of internal validity checks of data and, in a handful of cases, eliminated some institutions from certain analyses due to incomplete or irreconcilably inconsistent data. These cases are identified as they occur throughout the report. It should be noted that all financial aid data was <u>at the point of admission</u>. Although financial circumstances change over time -- other than adjusting all dollars into 2016 dollar

---

[22] College Board, “ACT and SAT Concordance Tables,” 2009, <https://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2009-40-act-sat-concordance-tables.pdf>.

[23] Only institution submitted a few ACT test scores from students post 2016. For this institution, the more recently adopted concordance table was used (though its validity has been disputed by ACT): College Board, “New SAT to ACT Concordance Table” 2016 <https://collegereadiness.collegeboard.org/pdf/higher-ed-brief-sat-concordance.pdf>

[24] EFC, Gift Aid, and Total Cost of Attendance were adjusted using the St. Louis Fed's FRED II database urban consumer CPI (CPIAUCSL), base year 2016.

equivalents, we did not attempt to track any changes over time. Thus, for instance, when we discuss graduation rates or college GPAs in the context of financial need, it is based on the financial need or aid award at the time of admission.

Similar to HSGPAs above, we encountered a number of challenges in interpreting financial aid data, particularly with respect to merit aid. In some instances, the current staff at the institution were unable to affirm the protocols for recording certain financial aid data elements that were used in the earlier years. In some cases, merit aid was only defined as gift aid awarded to no-need students, whereas at other institutions, virtually every student receiving need-based aid has a portion of it labeled as a "merit" scholarship. Some institutions reported having no institutional merit scholarship program, and yet had a number of students flagged as merit scholarship recipients (leaving us to assume these may refer to externally awarded merit scholarships or to scholarships such as National Merit or Posse Foundation that are awarded by the institution, but to an externally-determined population). Thus we created a data element that identified students who received gift aid (from any source) that was in excess of their total demonstrated financial need. We primarily used this more consistent definition in place of institutional definitions of merit.

Ten institutions were able to provide information about financial aid awards made to admitted and enrolled students, five were able to provide reliable data only for students who enrolled, and six were unable to provide any financial award data. Pell grant recipients were identified by all but two institutions at the enrollee level, but less than half of them were also able to provide it at the admit level.

While conducting our analysis, we identified a number of students with apparent incongruities in the four financial data elements submitted. Two institutions had enough incongruities that we were compelled to exclude them from the analyses of financial aid. One significant incongruity didn't require exclusion of the data, but caused us to reconsider our interpretation of it. Specifically, of about 24,750 admits during the TOP cohort years identified as Pell recipients, we found slightly under 1,500 that had EFCs that were higher than would be Pell-eligible. Some of these records were actually No-Need, and the size of the Total Gift awarded appeared to be solely the Pell Grant. Such circumstances can occur when the federal EFC qualifies a student for a Pell Grant, but the institutionally-determined EFC takes into consideration additional resources, such as the income of the non-custodial parent. While most public institutions accept the federal EFC, many private institutions consider the additional resources when determining eligibility for institutional aid. We made the decision to assume that the records identified as Pell recipients were accurate, regardless of the EFC or Total Gift Aid reported, but it gave us a new appreciation for the potential nuances related to using "Pell recipients" as a proxy for low SES students.

The other major incongruity we discovered among our financial aid data points that we were unable to confidently decipher was the group of students who reportedly received gift aid in excess of the Total Cost of Attendance (COA) at their institution. These discrepancies ranged from a few dollars to many thousands of dollars. While we speculated that there were likely some students who received such extraordinary scholarships or combinations of scholarships that their total COA was indeed exceeded by their gift aid, the inclusion of these students had the potential to inappropriately distort some of our comparisons, so we have eliminated from our financial aid comparisons all students whose reported Total Gift Aid exceeded their COA by more than $1,000. A total of 927 students from 11 institutions (from a total of 765,087 records with reliable FA data) have been excluded from the FA comparisons for this reason.

Because of the varying costs at the participating institutions, rather than conduct our analyses based upon just the EFCs of the students, we instead focused on their ability to contribute to the institutions to which they had applied by assigning them to one of the four student segments outlined below. These segments are designed to be institutionally-specific. It is possible that if a student applied to more than one of the institutions in our study, and those institutions had substantially different COAs, the student may have been assigned to different segments for each of the institutions. This segmentation attempts to account for the financial circumstances of the family in direct relation to the costs of a particular institution, as well as to track institutional awarding strategies that are keyed to the institution's own costs.

The following segments were calculated by subtracting the adjusted EFC from the Total Cost of Attendance:

High Demonstrated Need
(HNEED)
Students who need 75%-100% of Total Cost of Attendance

Moderate Demonstrated Need
(MNEED)
Students who need 25%-74% of Total Cost of Attendance

Low Demonstrated Need
(LNEED)
Students who need 2%-24% of Total Cost of Attendance

No Need
(NNEED)
Students who did not request financial aid, or needed less than 2% of Total Cost of Attendance.

We then developed a systematic way of viewing student financial support in five "Need Met With Grant" Segments. The reader should note that for this purpose, we are evaluating whether need was met solely with gift aid. Traditionally, an institution "meets need" with a combination of gift aid (grants and scholarships), loans, and work-study – so, many of the high-need students (and even moderate need students) will fall into the category of "Need Not Met with Grant." In some instances these students may have been "gapped," but the reader also should not conclude that students in this category didn't have their need met under the more traditional definition.

Need Not Met with Gift
Gift aid less than demonstrated need by at least $1,000 (Need-GiftAid >1000)

Need Met with Gift
Gift aid roughly equal to need (Need-Gift Aid is between -1000 and +1000)

Need Exceeded with Gift
Gift aid exceeds demonstrated need by at least $1,000 (Gift Aid-Need > 1,000)

No Need, No Gift
No demonstrated need, and no gift aid awarded.

No Need + Gift
No demonstrated need, and gift aid awarded.

Also, please note that students flagged as "No Need" (with or without gift aid) includes not only those whose EFC was greater than the COA, but also those for whom the college did not record an EFC. In some instances, this may represent a student who simply never completed the Need Analysis process at that college (they already enrolled elsewhere, or had received a large enough scholarship that they didn't need to be considered for need-based aid.) So this may mildly overstate the number of true No Need students.

Finally, to assist in our analysis, we created an attribute called "Family Financial Contribution," or FFC. This was used to determine the total amount the family was expected to pay at a specific institution, and was calculated as:

Family Financial Contribution = Total Cost of Attendance - Total Gift Aid

Although the way in which each family covered their FFC is unknown (i.e., family resources, loans, work-study, additional outside scholarships, or assistance from relatives or friends), this number provides a more consistent basis for comparison than the EFC, as it represents the entire amount the family needed to provide exclusive of the gift aid provided in the institution's aid award.

Academic Outcome Data

We requested the critical academic measures of college success, along with a few additional areas of interest. We collected: the first-year college GPA (FYGPA), most recent (or final) CumGPA, major designation, a current student enrollment status, and an up-to-date graduation status for all students who enrolled. We used graduation status as our ultimate measure of student academic success.

The data was collected in 2016, so for a cohort that enrolled in 2008, the graduation rate would reflect an eight-year graduation rate, whereas for a 2012 cohort, it would reflect only a four-year graduation rate. The graduation rate data for TOP policy years will focus on the 4- and 5-year graduation rates of student cohorts entering in 2012 and 2011, respectively. Our participants adopted the policy across a range of timespans, so we have a subset of only 12 that had the policy in effect long enough to show both of these rates.

The FYGPA is the most consistent and consistently available college academic indicator, as it was reported at the end of the first year of college regardless of which cohort the student was in. The Cum GPA represents the last recorded GPA at the institution. So, for students who graduated, it will represent their cumulative GPA at graduation, but for students who have not yet graduated or who have left the institution, it represents their final or most recent cumulative GPA, which may represent anything from one year to several years' worth of academic work.

The college academic performance data was used primarily to assess any potential differences between the performance outcomes of the Non-Submitters and the Submitters. To a lesser extent, we also attempted to explore whether there was any significant difference between the overall performance of the pre-TOP cohorts and the post-TOP cohorts at each institution, but it is difficult to isolate the effect of the TOP from the effects of other changes that may have been occurring at the institution.

Peer Data

For another stage of the study we attempted to assess how changes in the size and composition of the applicant pools for the participating TOP institutions compared with those of a matched sample of their competitive-overlap peer institutions. We asked each TOP institution to identify three institutions they viewed as "peer competitor" institutions – not aspirational institutions, but institutions with which they have large applicant overlap and have a roughly even "win-loss" record for admitted students. For that portion of the study, to ensure comparability of data, we collected IPEDS data for both the TOP institutions and their Test Required Peers for each of the matched cohorts of students. (For more detail on the selection of Peer institutions, see page 19.)

The participating institutions in the study were given a data request of approximately 40 variables. The following text describes the data definition.

I. Data type: Record-level data on students at the point of application and additional data on those who matriculated.
II. Population definition at point of application or entry to college: First-time, undergraduate, bachelor's degree seeking, full-time entering in the fall. Please exclude: transfer-in students, graduate students, part-time, non-degree seeking, associate's degree seeking, unclassified students, spring or summer entry students.
III. Years of data requested: In total, four cohorts requested based on the start date of the TOP policy on standardized testing in the admissions decision.
    a. Two cohort years ***leading up*** to the adoption of the test-optional policy.
    b. Two cohort years ***post*** adoption.

The following variables were requested of each participating institution:	
StudentID	Unique, blind
Cohort	Year of entry to college, full time, first time, fall entry students.
FirstContact	First point of contact between the student and the institution.
FirstContactDate	Date of the first contact.
AppDate	Date of the application receipt.
AppType	1 = Early Action or Early Decision 0 = Regular
ScoreConsid	Did the student apply with or without their standardized test score being considered as part of the Admissions decision? 1 = Test Score Considered (SAT I OR ACT reviewed in the admissions decision) 0 = Test Score Not Considered (No testing reviewed in the admissions decision) 2 = Alternative Test Considered (International Baccalaureate, TOEFL, or British A-Levels were submitted instead of SAT I, ACT)
AdmitStat	1 = Admitted (Include students who were pulled from the wait list and admitted) 0 = Denied/Wait List 2 = No Decision Rendered (Student withdrew prior to decision, or application incomplete)
StuDec	1 = Student Enrolled 0 = Student Did Not Enroll

	(Student Declined Offer of Admission or Withdrew Before Enrollment) 2 = Deferred
Zip	Student hometown five-digit zip code from admission record.
Gender	1 = Female 0 = Male 2 = Other
RaceEth_Instit	New IPEDS codes used for the cohorts after the IPEDS code change. Nonresident Alien Hispanic/Latino American Indian or Alaska Native Asian Black or African-American Native Hawaiian or Other Pacific Islander White Two or more races Race and ethnicity unknown
BirthYear	Birth year in 4 characters (YYYY).
CEEB	Student's graduating high school CEEB code.
HSType	Student's graduating high school type.
HSGPA	Cumulative High School GPA
HSGPAScale	Institution record of high school GPA.
SATCR	Score submitted for admission, scale 200-800
SATMath	Score submitted for admission, scale 200-800
SATWriting	Score submitted for admission, scale 200-800
ACTComp	Score submitted for admission, 2 digits
EFC	EFC used to award aid at the time of admission (from any formula: institutional, federal, etc.), dollar amount - 0 means no contribution - Null is no FAFSA submitted (no need) - 1.00 (or other single digit) means $1 required contribution

Pell	Did the student receive Pell grant aid at the time of admission? 1 or 0
NeedAid	Need-based grants received at the time of admission. 1 or 0 (Need-based scholarship or grant aid: Scholarships and grants from institutional, state, federal, or other sources for which a student must have financial need to qualify.)
MeritAid	Institutional merit scholarships/grants at the time of admission. 1 or 0
TGiftAid	Total packaged grant for all admitted students in their first cohort year, including any combined institutional, governmental and private aid.
PEAFather	1 = Middle school/Jr. high 2 = High school 3 = College or beyond 4 = Other/unknown
PEAMother	1 = Middle school/Jr. high 2 = High school 3 = College or beyond 4 = Other/unknown
FirstGen	Institutional record of "neither parent having received a four-year college degree"
Athlete	Recruited athlete in the admissions process.
EnrollStat	Current status as of Spring Term 2016 1 = Active (Please include institution-approved off-campus study) 2 = Leave of absence 3 = Withdrew before graduating (If recorded, please exclude known transfers and code them as “6”) 4 = Graduated 5 = Deceased 6 = Transferred-out 7 = Unknown 8 = Did Not Enroll
GradYear	Actual graduation year, if graduated, as of Spring Term 2016.
PrimaryCIP	Major code, Primary CIP, Undecided
SecondCIP	Second major code
FYGPA	Cumulative GPA at end of the first year (June), 0 – 4.0 point scale. (Note: If student withdrew before end of year, ending GPA)
CumGPA	Final Recorded Cumulative GPA on a 0 – 4.0 point scale. (Note: If student withdrew or graduated, ending GPA)

LD	Learning difference or learning disability, submit institutional codes and code definitions.
AR	The overall academic rating as decided by admissions staff during the application evaluation process, if your institution records such a summary code. If a holistic admission rating is used, please also include in a separate column.
PR	The overall personal rating as decided by admissions staff during the application evaluation process, if your institution records such a summary code.
AltEval	Include any alternative means of student evaluation, particularly any that were added at the time of testing de-emphasis. For example, interview rating scale, questionnaire score, portfolio rating, strength of curriculum, etc.
ESL	First or primary language is not English. (Common Application question regarding "First Language")
The following variables were transformed from the above list into those used in the analysis:	
FirstContact_Universal	Universal categories of first contact created for the study.
URM	1 = Hispanic/Latino, American Indian or Alaska Native, Black or African American, Native Hawaiian or Other Pacific Islander 0 = White, Asian, NonRes Aliens, Two or More Races
Diversity	1 = Hispanic/Latino, American Indian or Alaska Native, Black or African American, Native Hawaiian or Other Pacific Islander, Pell, First Gen 0 = White (all remaining cells), Non-Res Aliens, Two or More Races
Expanded Diversity	URM Only, URM+Pell, URM+Pell+FG, URM+FG, No layers
HSGPA_ConvertAll	- Converted all known *unique* scales to 4.0 scale using scales provided by institution, or if missing, consulted the proprietary HS GPA scale database - Converted all 100-point scales to 4.0 - Allowed all others to remain up to 5.99
HSGPA_Truncate	Change all GPAs >4.0 to a flat 4.0
SATConvertAll	Combined SAT CR+M and ACTConv into one comprehensive list. (No writing scores) If both SAT and ACT scores submitted, highest score was used.
EFC-AdjCPI	Adjusted for inflation using the St. Louis Fed's FRED II database urban consumer CPI (CPIAUCSL), base year 2016.
COA	IPEDS total cost of attendance for entering cohort year, accurate for in-state and out-state residents (living on campus not with family), if different

TGiftAid_Adj	Adjusted for inflation using the St. Louis Fed's FRED II database urban consumer CPI (CPIAUCSL), base year 2016.
FFC (Family Financial Contribution)	COA-Adj - Total Gift Aid-Adj
DemonstratedNeed	COA-Adj - EFC-Adj No negative or 0 values
PercentNeed	(DemonstratedNeed/COA-Adj)*100
NeedSeg	Segments derived from value in PercentNeed: 1-HNEED: 75-100% 2-MNEED: 25-74% 3-LNEED: 2-24% 4-NONEED: =<2% (up to <$1000 DemonstratedNeed and NULL EFC)
QualifiedNeed	1 = NeedSeg 1-HNEED, 2-MNEED, or 3-LNEED 0 = NeedSeg 4-NONeed
NeedMet_Seg	Full Pay No Aid = NeedSeg 4-NONEED and TGiftAid-Adj NULL Full Pay + Aid = NeedSeg 4-NONEED and TGiftAid-Adj value Dem Need Not Met = Aid-Need$$ < -$1,000 Dem Need Met = Aid-Need$$ -$1,000 through +$1,000 Dem Need Exceed = Aid-Need$$ > +$1,000
FirstGen_Binary	Combine the three columns. 1 = PEAFather 1,2 AND PEAMother 1,2 OR FirstGen 1 0 = PEAFather 3 AND PEAMother 3 OR FirstGen_Instit 0 Null = PEA 4 (for both mother and father)
Top15	(Based on CEW) 1 Agriculture and Natural Resources 2 Arts 3 Biology and Life Science 4 Business 5 Communications & Journalism 6 Computers & Mathematics 7 Education 8 Engineering 9 Health 10 Humanities & Liberal Arts 11 Industrial Arts and Consumer Services 12 Law & Public Policy 13 Physical Sciences 14 Psychology and Social Work 15 Social Science

STEM	STEM Designated Degree Program List Effective May 10, 2016 1 = ScienceTechnologyEngineeringMath 0 = Non-Stem Major
GPAIncrease	1 = CumGPA higher than FYGPA 0 = CumGPA is lower than FYGPA
AR_10	Converted to 10 point scale, where 10 is the highest and 1 is the lowest
PR_10	Converted to 10 point scale, where 10 is the highest and 1 is the lowest

REFERENCES

Allman, M. B., Anci, D., Babington, C., Ballinger, P., Beseda, M., Bland, T., . . . Block, J., (2016). *Turning the Tide: Inspiring Concern for Others and the Common Good through College Admission*. Cambridge, MA: Harvard University. Retrieved from http://mcc.gse.harvard.edu/files/gse-mcc/files/20160120_mcc_ttt_execsummary_interactive.pdf

Belasco, A. S., Rosinger, K. O., & Hearn. J. C. (2014). The Test-Optional Movement at America's Selective Liberal Arts Colleges: A Boon for Equity or Something Else? *Educational Evaluation and Policy Analysis,* pp. 1-18.

Bowen, W, Chingos, M, McPherson, M. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press, 2009.

Buckley, Jack, Letukas, Lynn, and Wildavsky, Ben, Eds,. *Measuring Success: Testing, Grades and the Future of College Admissions*. Baltimore: Johns Hopkins University Press, 2018.

College Board, *New SAT to ACT Concordance Table* (2016). Retrieved from https://collegereadiness.collegeboard.org/pdf/higher-ed-brief-sat-concordance.pdf

College Board, *ACT and SAT Concordance Tables* (2009). Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-2009-40-act-sat-concordance-tables.pdf

FairTest List, (2016-2018) Retrieved from FairTest website https://www.FairTest.org/university/optional

Hiss, W, Doria, K (2011). Defining Promise: Twenty-five Years of Optional Testing at Bates College: 1984-2009. Paper presented at Conference, *Diamonds in the rough: Why Every Institution of Higher Learning Wants to Attract and Admit Dyslexic Students*. Stanford, CA: Stanford University Business School, June 4, 2011. http://www.bates.edu/admission/optional-testing/

Hiss, W, Franks, V. (2015). Brave New World: How Test-Optional Policies Could (and Should) Change College Admission. *The Journal of College Admission*, pp. 32-34.

Hiss, W, Franks, V. (2014). *Defining Promise: Optional Standardized Testing Policies in American College and University Admissions*. Retrieved from NACAC website:http://www.nacacnet.org/research/research-data/nacac-research/Documents/DefiningPromise.pdf.

Kobrin, J.L., Camera, W.J., and Milewski, G.B, (2002). *Students with Discrepant high school GPA and SAT I Scores*. Research notes. RN-15. New York: The College Board.

Lucido, J. (2018). [Understanding the Test-Optional Movement]. In Buckley, Jack, Letukas, Lynn, and Wildavsky, Ben, Eds,. *Measuring Success: Testing, Grades and the Future of College Admissions* (p. 164). Baltimore: Johns Hopkins University Press.

Mattern, K.D., and Patterson, B.F. (2014). *Synthesis of Recent SAT Validity Findings: Trend Data over Time and Cohorts*. New York: The College Board. https://files.eric.ed.gov/fulltext/ED556462.pdf

Mattern, K.D, Shaw, E.J., and Kobrin, J.L. (2010). *A Case for Not Going Test-Optional: Students with Discrepant SAT and HSGPA Performance*. http://files.eric.ed.gov/fulltext/ED563419.pdf

McNeish, D. M., Radunzel, J., Sanchez, E., (2015). *A Multidimensional Perspective of College Readiness: Relating Students and School Characteristics to Performance on the ACT*. ACT Inc.

Mulugetta, Y. M. (2015). *Understanding the Impacts of the Test Optional Admission Policy*. Retrieved from https://www.ithaca.edu/ir/docs/testoptionalpaper2.pdf

National Association for College Admission Counseling. (2016). *Use of Predictive Validity Studies to Inform Admission Practices.* Retrieved from https://www.nacacnet.org/globalassets/documents/publications/research/testvalidity.pdf

National Association for College Admission Counseling. (2008). Report of the Commission on the Use of Standardized Tests in Undergraduate Admission, Arlington VA NACAC, p.7, p. 11.

Sanchez, E., Edmunds, A. (2015). *How discrepant high school GPAs, test scores, and engagement impact persistence.* Presentation at the annual convention of the Association for Institutional Research, Denver, CO.

Sanchez, E., Mattern, K. (2018) [When High School Grade Point Average and Test Scores Disagree]. In Buckley, Jack, Letukas, Lynn, and Wildavsky, Ben, Eds,. *Measuring Success: Testing, Grades and the Future of College Admissions* (pp. 118-141). Baltimore: Johns Hopkins University Press.

Soares, J. A. (Ed.) (2012). *SAT: Wars: The Case for Test-Optional College Admission.* New York, NY: Teachers College Press.

Syverson, S., (2007). *The Role of Standardized Tests in College Admissions: Test Optional Admission*. In T. Crady & J. Sumner (Eds.), *Key Issues in New Student Enrollment: New Directions for Student Services* (no.118., pp. 55-71). Wiley Periodicals, Inc.

The Test and the Art of Thinking Davis, Michael. Canobie Films, 2018. Film. http://thetestdoc.org/