Early Findings from the Experimental Evaluation of the Undergraduate Research Scholars Program at the University of Wisconsin-Madison

A Wisconsin HOPE Lab Research Brief

David B. Monaghan, Emily Colo, & Sara Goldrick-Rab
June 2017

Executive Summary

The Undergraduate Research Scholars (URS) program at the University of Wisconsin-Madison was created to allow undergraduates to directly participate in scholarly inquiry, increase participation by students from under-represented groups in academic endeavors, and improve retention and college completion of students from under-represented groups. The program has two components. First, participants work as a research assistant for an on-campus researcher. Second, they take a support seminar led by former program participants.

To learn about impacts and refine implementation, the URS program in 2014 initiated a randomized control trial evaluation conducted by the Wisconsin HOPE Lab. In this brief we present one-semester post-treatment outcomes for the 2015-16 URS applicant cohort, with analysis restricted to outcomes measurable through administrative data and relevant to URS' goal of improving retention among underserved students. Though average impacts for the full sample were negligible, URS appears to increase credit accumulation for first-generation college-goers, and particularly for minority first-generation students.

The Wisconsin HOPE Lab will continue to follow the progress of this and the following two cohorts' progress for another year. In spring 2018 we will produce an evaluation report synthesizing experimental results with insights gleaned from interviews and focus groups conducted with program participants.

Introduction

With wages and working conditions deteriorating for those with a high school diploma or less, a college degree is more consequential than ever (Abel & Deitz 2014). Yet despite increasing enrollment, class- and race-based disparities in college completion have grown in recent decades (Bailey & Dynarski 2011). This pattern extends to highly selective colleges, where nationally only 61% of Black and 59% of Latino students complete a bachelor's degree in six years, compared with 80% of white and Asian students. Similarly, the graduation rate gap between first-generation college-goers and the children of college graduates is more than 15 percentage points (83% vs. 67%) (NCES 2011; authors' calculations).

The University of Wisconsin-Madison is no exception. In recent cohorts, 87% of White students completed a bachelor's degree within six years, compared with 78% of Latino, 75% of Black, 70% of Southeast Asian, and 64% of Native students. And first-generation college-goers were 8 percentage points less likely to graduate than those whose parents hold a bachelor's degree (UW-Madison Academic Planning and Institutional Research 2016; authors' calculations).

Colleges have created numerous programs to address these attainment gaps. These include "college skills" classes, designated support personnel, freshman orientation programs, learning and living/learning communities, and peer mentorship. Another such initiative is UW-Madison's Undergraduate Research Scholars (URS) program. URS has among its goals improving retention of under-represented minority and first-generation college students through scholarly socialization and peer support. While available to all students, URS prioritizes recruitment of minority and first-generation students.

In what follows, we discuss early results from an experimental evaluation of this program. For reasons of data availability and policy relevance, we limit investigation to program effects on major indicators of academic progress: retention, grade point average (GPA), and credit accumulation.

The URS Program

Researchers accounting for class- and race-based disparities in college completion point first to economic strains and differences in academic preparation (Adelman 2006; Goldrick-Rab 2016). Residual differences in completion are commonly accounted for in one of three ways. The first foregrounds disparities in knowledge of how college "works" and of how to "be" a college student. Such implicit understandings are gleaned "naturally" through middle-class socialization, and colleges presume them in students. Students from working-class families may not have internalized such knowledge, leading to consequential errors which harm their progress (Nora 2004). The second explanation highlights the importance of "integration" into the collegiate social world (Tinto 1987). For various reasons – cultural differences, prejudice, a lack of resources, etc. – students from disadvantaged backgrounds may be less inclined or able to establish supportive social

relationships on campuses in which white, middle-class culture is dominant (Kuh & Love 2000). Without this support, they are less likely to persist. The final explanation underlines internalization of and identification with the academic mission of the university. An interest in one's program and subject matter in itself rather than for instrumental reasons can make it easier to perform well and persist, but it is thought that such a predisposition towards academics is less likely to be inculcated by less-educated parents (Arnold & Doctoroff 2003; Walker *et al*. 2006).

The URS program aims to reduce disparities by class and race by addressing all three of these hypothesized mechanisms. It seeks to infuse college knowledge by exposure to knowledgeable faculty and upper-level student mentors. It provides opportunities for the formation of supportive on-campus relationships with these mentors and with peers who have similar interests. And finally, through intensive, individualized, experiential learning, it attempts to develop students' identities as scholars.

Created in 1998, URS is a two-semester credit-bearing course in which participants ("URS scholars") work as assistants to mentors engaged in original research or creative enterprise, and also attend a weekly one-hour seminar led by a pair of upper-level undergraduates ("URS Fellows"). Typically the mentor is a faculty member, but postdoctoral researchers, research scientists, and graduate students are also eligible. The scholar-mentor pairing is accomplished mutually: scholars select a mentor from a list but must then interview for the position. Though the program provides guidelines, specific arrangements— responsibilities, hours, etc. — are developed by mentors and formalized in "Research Contracts" which scholars agree to by signing. The weekly seminar covers research design, critical reading and writing, and practice in presenting research or creative work to a broad academic audience. Each year, URS leaders develop a central theme for the seminars, but week-to-week curricular planning is left to fellows. Toward the end of the spring semester, scholars present results of their research or creative activities at an annual symposium organized by the Office of the Provost and open to the public. Students receive a grade based on assessment by the research or creative work mentor (75%) and by the fellows conducting their seminar (25%).

URS is open to all freshmen, sophomores, and new transfer students, but has relationships with minority- and first-generation-serving programs such as the Pre-College Enrollment Opportunity Program for Learning Experience (PEOPLE), the Posse Foundation, the Center for Academic Excellence, the Center for Educational Opportunity, and the Chancellor's/Powers Knapp Scholarship Program to ensure an over-representation of such students. In turn, some of these programs require participation in a "high impact practice" program such as URS. Remaining participants are recruited from the undergraduate population at large through word-of-mouth and recruitment at orientation events.

URS has five official goals:

1) to provide opportunities for early undergraduates to experience scholarly inquiry in research or creative work
2) to create a scholarly community inclusive of people from historically underrepresented groups
3) to improve retention and nurture aspirations of students, including those in historically underrepresented groups
4) to develop leadership through peer mentoring
5) to enhance the University's mission of research and creative endeavor

This report focuses only on the third goal. As mentioned above, there are three mechanisms through which URS is thought to further this goal. The program helps to infuse college knowledge through providing the student with college-savvy mentors who can answer their questions and offer advice. Both the research mentor and the upper-level fellows can serve these roles. Additionally, the seminar provides explicit instruction in certain college and career skills. URS also provides opportunities for forming supportive social relationships on campus. It does so first of all by pairing the student with a research mentor who can provide assistance, connections to further opportunities, and support. Equally important, the informal atmosphere of the seminars, which encourage lively participation and self-expression, is conducive to helping students open up to each other and spark friendships. Finally, the program develops students' identities as scholars in a number of ways. Research mentorships allow students to participate actively in the creation of new knowledge, helping them to understand the open-ended and creative nature of inquiry. Students also make connections between what they are learning in their mentorships and other course material. Through the seminars they grasp practical similarities and differences between disciplines, learn the academic practices of understanding the views of others, assessing their own, and questioning ideas they had taken for granted. Through these experiences, it is hoped that they gain confidence in themselves and their academic abilities, and come to understand themselves as having a realistic potential future in research or creative work.

Undergraduate research experience programs are common in postsecondary education and have received a fair amount of scholarly attention. Research has associated participation in such programs with gains in intellectual development and career clarity (Baur & Bennett 2003; Hunter *et al*. 2007; Lopatto 2004, 2007; Seymour *et al*. 2004), as well gains in college completion and graduate school enrollment (Jones *et al*. 2010; Lopatto 2007; Zydney *et al*. 2002). Some studies find larger effects for minority students (Ishiyama 2007; Jones *et al*. 2010; Lopatto 2007). However, most of these programs lack an equivalent of URS's fellow-led seminar. And most of this research is observational and so cannot disentangle program impacts from factors influencing participation.

There are two exceptions, however. In the first, the Undergraduate Research Opportunity Program (UROP) at the University of Michigan, Ann Arbor, participants also work with a faculty researcher, take part in peer-led "research interest groups", and present research at on-campus symposiums.

In a randomized control trial evaluation, researchers found that UROP participation increased persistence and graduation among African-American students and those who had lower high school GPAs, and also raised the rate of graduate school attendance (Gregerman *et al*. 1998; Hathaway *et al*. 2002).

The second is the federally-sponsored Research Initiative for Scientific Enhancement (RISE) program. Better-resourced than either URS or UROP, RISE involves paid research positions which last all four years of college, a seminar focused exclusively on research methods, expenses-paid summer research programs, and graduate school application support. Using propensity-score methods in a sample of minority students at 25 institutions, Schutz and colleagues (2011) find that RISE participation increased persistence in STEM intentions

Data and Methods

Programs often select participants according to characteristics predictive of success. Thus, naïve comparison of participants to non-participants is likely to be misleading, as apparent impacts may in fact be attributable to pre-existing characteristics (i.e., to selection bias). The use of a lottery to select among eligible applicants renders participation independent of background characteristics, and is appropriate when demand for a program exceeds supply. In the case of URS, 300-350 students apply annually for 180 available slots.

In consultation with the Wisconsin HOPE Lab, the program admitted applicants through a lottery in the 2015-16 and 2016-17 academic years. To remain faithful to its mission during evaluation, URS leaders reserved slots (47 in 2015, 42 in 2016) for students from underserved populations. These applicants were directly admitted rather than being placed into the randomization pool, and are excluded from experimental estimates. Remaining applicants were randomly assigned, in batches and on a rolling basis, to either treated or control groups at a 2:1 ratio.

We obtained baseline socioeconomic, academic, and demographic measures, as well as measures of academic performance in college, from the Registrar and the Office of Financial Aid. We focus on three outcomes: retention, GPA, and credits earned. While credits earned is a function of both credits attempted and a student's success in their courses, in this sample only four out of 261 retained students completed fewer than 100% of their attempted credits, rendering separate measures redundant.

Assessing Baseline Equivalence

Because random assignment occurred on a rolling basis, and because administrative data was not available until much later, baseline equivalence could only be assessed post-hoc. In keeping with recommendations of the What Works Clearinghouse (WWC) (2014), we assess equivalence in terms of both t-tests of differences in means, and effect sizes. The latter express differences

in means as a proportion of a standard deviation – larger effect sizes indicate greater divergence between treatment and control groups. In experimental studies with minimal attrition, WWC does not set benchmarks for "acceptable" effect sizes, since randomization eliminates selection bias. We include them because they permit comparisons among variables.[1]

Table 1 presents characteristics of the 2015-16 applicant cohort. As described above, URS seeks to reduce class- and race-based gaps in academic outcomes. As a result, relative to the university's undergraduate population, racial minorities are over-represented in both the applicant and randomization pools. At UW-Madison in 2015, 2% of undergraduates were Black, 4.7% were Latino, and 5.5% were Asian, compared with 11%, 19%, and 21% of URS applicants. In fact, given that only 629 entering freshmen in 2015 were "targeted minorities",[2] in 2015 URS enrolled the equivalent of 18% of all such students.[3] URS applicants are also more likely to be first-generation college-goers (34% vs. 18%) and to be recipients of federal need-based aid (34% vs. 14%) than UW's undergraduates as a whole (UW-Madison Academic Planning and Institutional Research 2017; authors' calculations). The over-representation of disadvantaged students is more pronounced among direct-admit students, and reflects the targeted recruiting practices of the URS program in advancing its goals of diversifying participation in research and improving retention of underrepresented groups.

[1] For continuous outcomes, the recommended effect size metric is Hedges' *g:*

$$g = \frac{\omega(y_t - y_c)}{\sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{n_t + n_c - 2}}}$$

In the above, *t* and *c* denote the treated and control groups, *y* is the mean of the variable in question, *n* is the group sample size, *s* is the group standard deviation, and *ω* is a small sample correction defined below. For dichotomous variables, the comparable metric is the Cox index:

$$d_{Cox} = \omega[\ln\left(\frac{p_t}{1 - p_t}\right) - \ln(\frac{p_c}{1 - p_c})]/1.65$$

where *p* is the group probability for the variable being assessed. The small sample correction factor is given by

$$\omega = 1 - \frac{3}{(4N - 9)}$$

Where *N* is the full sample population.

[2] At UW-Madison, this includes Black, Latino, Native American, and Southeast Asian students.

[3] We write "the equivalent" because URS is open to and enrolls both freshmen and sophomores

Table 1: Socio-demographic and academic background characteristics of 2015 URS applicants

	All students (N=270)		Direct admits (N=47)		Treated (N=149)		Control (N=74)		Treat vs. control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*p*	Effect size
1st generation college	0.34		0.51		0.32		0.27		0.491	0.13
White	0.39		0.09		0.48		0.42		0.419	0.14
Asian	0.21		0.23		0.20		0.20		0.981	0.01
Latino	0.19		0.36		0.14		0.19		0.353	0.21
Black	0.11		0.17		0.09		0.14		0.270	0.30
Native	0.04		0.09		0.04		0.03		0.619	0.25
Female	0.54		0.64		0.54		0.46		0.278	0.19
ACT	28.21	3.78	27.35	3.59	28.37	3.96	28.51	3.48	0.798	0.01
Freshman	0.59		0.38		0.60		0.68		0.297	0.21
In-state	0.58		0.57		0.63		0.49		0.040	0.36
US citizen	0.92		0.89		0.93		0.92		0.705	0.12

The admission pool also reflects the highly selected character of the overall UW-Madison population. Most of the applicants – including almost half of directly admitted students – have college-educated parents. The average ACT score is over 28 (similar to 28.7 for the entering class of 2015) and is only slightly lower for directly-admitted students. This means that the majority of URS applicants – including direct-admits – scored in the top 15% of the national ACT distribution.

Treatment and control groups are similar in terms of background characteristics, indicating that randomization succeeded in rendering treatment status independent of pre-existing characteristics. However, because of small sample sizes and chance randomization error, there are some differences. For instance, the difference between treated and control groups in the proportion from Wisconsin is statistically significant at $p<.05$, and is equivalent to over a third of a standard deviation. As described below, we control for these background measures in our analyses.

Estimating Treatment Effects

We report intent-to-treat (ITT) estimates, which compare groups defined through treatment assignment, without an adjustment for compliance. Students were randomized before they were assigned to groups, but background characteristics might still play a role in whether or not they decided to actually enroll in URS. ITT estimates are unaffected by this selective compliance and are better causal estimates than comparisons of program participants to non-participants. As enrolling in URS was impossible without program permission, compliance was imperfect only for those randomized to treatment. In the 2015-16 cohort, 20.8% of those assigned to treatment (31 students) did not enroll in URS. Analysis suggests that compliance was slightly better among first-generation college students relative to continuing-generation students (89% vs. 75%), and out-of-state students relative to Wisconsin residents (83% vs. 77%) (see Appendix). We produced estimates which take compliance into account (complier average causal effects; Angrist *et al*. 1996); results were similar to those obtained through ITT analyses.

We assess experimental outcomes through both raw group differences and through regression adjustment, which reduces the influence of randomly-occurring group differences on estimates. We calculate treatment effects in both native units and in effect sizes (as defined above). WWC suggests that effect sizes of 0.25 merit consideration as “substantively important”.

Findings

In Table 2, we present mean outcomes separately by admit/treatment status for the 2015-16 applicant cohort during the first post-treatment semester (fall 2016). The top panel shows the full applicant pool. Here we draw attention to the very strong academic performance of these students. The control group was retained at a rate of 96%, earned nearly 15 credits on average, and had a mean GPA of 3.3. This high baseline performance means that for the full sample of applicants, sizeable program impacts are unlikely.

Table 2: Descriptive one-semester post-treatment outcomes for the 2015-16 URS applicant cohort

	Treated	Control	N
Full sample			
Retention	0.96	0.96	223
Credits earned	14.73	14.71	218
GPA	3.35	3.31	218
Under-represented minority			
Retention	0.92	0.97	116
Credits earned	14.00	14.07	112
GPA	3.04	3.12	111
First-generation			
Retention	0.96	0.90	92
Credits earned	14.73	13.37	86
GPA	3.11	3.01	85
Minority and first-generation			
Retention	0.97	0.93	58
Credits earned	14.53	12.8	55
GPA	2.98	2.99	55

A more appropriate test group would seem to be minority students. But among targeted minorities in the URS program, only half are first-generation college goers. The second panel shows that 97% of control-group minority students are retained, and their academic performance is very strong. Among first-generation college-goers, only 90% of control group students were retained, and these students earned 13.4 credits and a GPA of just over 3.0 on average. In this population, then, a program could produce appreciable gains. This is even truer among minority first-generation students. The bottom panel shows that control students from this population earned 12.8 credits and a 2.99 GPA on average.

Table 3 displays program impacts both with and without regression adjustments. In the full sample there are no appreciable differences between treated and control groups in terms of academic outcomes. This, of course, is to be expected: the goal of the URS program is to improve outcomes specifically among disadvantaged groups rather than among students generally. Similar experimental results are obtained for targeted minorities, which is surprising given the program's mission but not given these students' strong baseline performance.

Table 3: First post-treatment semester experimental outcomes for URS 2015 cohort

	Unadjusted treatment effect	Regression-adjusted treatment effect	p-value	Effect size
All students				
Retention	0.00	0.01	0.879	0.18
Credits earned	0.02	0.12	0.721	0.02
GPA	0.04	0.04	0.542	0.14
Under-represented minority				
Retention	-0.04	-0.01	0.862	0.18
Credits earned	-0.06	0.42	0.545	0.05
GPA	-0.08	-0.04	0.783	0.12
First-generation				
Retention	0.06	0.04	0.633	0.33
Credits earned	1.36	1.15	0.096	0.21
GPA	0.10	0.10	0.558	0.28
Minority and first-generation				
Retention	0.03	0.04	0.606	0.53
Credits earned	1.73	1.86	0.047	0.27
GPA	-0.01	-0.05	0.804	0.14

Among first-generation students, point estimates suggest that program participation increased retention by 4 percentage points and increased GPA by over a quarter of a standard deviation. Significance tests show that these impacts are quite possibly the result of randomization error, so we urge caution in interpreting this results. There is stronger evidence of program impacts on credit accumulation, with treated students earning over a credit and a third more in the first post-treatment semester than control peers.

Finally, among students who are both first-generation and targeted minorities, retention and GPA effects are small, but URS participants earned nearly two more credits on average than their control-group peers. We emphasize that this indicates taking a higher credit load, not differences in course completion. It may not, however, mean that program students are taking more difficult courses.

Table 4: Interaction models showing variation in treatment effects on credits earned in first post-treatment semester by student subgroup

	Model 1	Model 2	Model 3
Treated	0.0590	-0.322	-0.276
	(0.441)	(0.405)	(0.382)
Targeted minority	-0.887	-0.734*	
	(0.622)	(0.423)	
First-generation	-0.0414	-1.154*	
	(0.409)	(0.675)	
Minority and first-generation			-1.946***
			(0.730)
Treated X minority	0.165		
	(0.706)		
Treated X first-generation		1.572**	
		(0.764)	
Treated X minority first-gen			2.097**
			(0.843)
Controls	Y	Y	Y
Constant	14.52***	14.93***	14.06***
	(1.712)	(1.707)	(1.537)
Observations	218	218	218
R-squared	0.114	0.131	0.130

*** p<0.01, ** p<0.05, * p<0.1

To display program impacts on credit-completion another way, we examine the variation in effects by student subgroup in Table 4. The interaction terms in these three models tell us how much larger (or smaller) the average treatment effect is for minority, first-generation, and minority first generation students (respectively) compared to other students. Again, there does not appear to be any substantial difference in effect size by minority status. But the program's impact on first-generation students is 1.6 credits larger than on continuing-generation students. And its impact on first-generation minority students is over two credits larger than the impact on non-first generation students. These results are both statistically significant at $p<.05$, constituting strong evidence that the program has larger impacts on those it is designed to serve.

Conclusion

Because experimental evaluations are rare in higher education, we cannot often precisely estimate causal effects of student-service programs. Without such studies, administrators are left to tease out whether programs actually meet the goals with which they are charged on the basis of observational data. This evaluation thus represents a rare opportunity.

URS seems to have its largest short-term impacts on credit accumulation among first-generation college-goers. For all first-generation students, URS seems to have raised credit completion by 1.36 credits, with even larger gains for first-generation minority students. Point estimates also suggest gains in retention, but we await further data collection to address this question with greater certitude.

That no appreciable gains are observed for the full sample of students is to be expected given the very strong academic preparation of UW-Madison students generally. In fact, the pattern of null effects on the full sample and stronger effects on the targeted sub-population was also observed in the UROP evaluation at the University of Michigan. The URS program is formally open to all students, and obtains over-representation of minorities and first-generation students only through targeted recruitment and reservation of places. Given the predominantly white and middle-class profile of the university's student body, it is simply not possible to fill the program solely with students from targeted populations. The remainder of places are therefore filled by high-achieving students from college-educated families. While these students hope to, and most likely do, benefit from taking part in URS, it will not be in terms of retention, grades and credit-taking. The benefits such students obtain are more likely to be clarifying career goals, obtaining practical understanding of research processes, and intellectual development – none of which we can measure here.

One expects to see academic gains in URS's target population of first-generation and targeted minority students, among whom there are some early indications of program impacts. It should also be recalled that the URS program directly admitted forty-students who best exemplify its "target population", and that these were removed from the evaluation pool. As these are precisely the students the program was designed to serve, it is among them that impacts ought to be

largest. But no plausible counterfactual cases exist through whom we can determine whether and how the program impacted these students. Because it is possible that the URS program more strongly impacted directly admitted students than those for whom it was possible to randomize exposure, this raises the matter of what economists James Heckman and Jeffrey Smith (1995) call *randomization bias.* Randomization bias is a concern about external validity, in which estimates obtained in a population which *could be recruited* for an experiment serve as proxy for what likely occurs in a population that a program *typically serves.* If these two populations differ systematically in terms of response to the treatment, then such a generalization leads to incorrect conclusions regarding program impacts. In this case, we suspect that such bias, if it exists, would lead us to *underestimate* treatment impacts rather than to overestimate them, because we are testing the effectiveness of the program on less disadvantaged students than it is designed to serve.

With only one semester of outcomes available, and for only one cohort, current findings are provisional. In the fall of 2017, URS will randomize a third and final cohort of applicants, bringing the total experimental pool to over 600 students. Since most new students will be freshmen, they will not be expected to graduate until 2021, and final results of this experiment will not be available until even later. And ultimately what is most important in this study is the program's longer-term effects on retention and completion. Present results suggest that the URS is having a small, but appreciable impact on the groups it is designed to serve. Given the stubbornness of race- and class-based achievement gaps, this is good news indeed.

Sources

Abel, J.R. & Deitz, R. (2014). Do the benefits of college still outweigh the costs? *Current issues in economics and finance, 20*(3):1-11.

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, D.C.: US Department of Education.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444-455.

Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology, 54*(1), 517-545.

Bailey, M. J., & Dynarski, S. M. (2011). Inequality in postsecondary education. In G.J. Duncan and R.J. Murnane (Eds.), Whither opportunity? *Rising inequality, schools, and children's life chances* (pp. 117-132). New York: Russell Sage.

Bauer, K. W., & Bennett, J. S. (2003). Alumni perceptions used to assess undergraduate research experience. *The Journal of Higher Education, 74*(2), 210-230.

Goldrick-Rab, S. (2016). *Paying the price: College costs, financial aid, and the betrayal of the American dream.* Chicago: University of Chicago Press.

Gregerman, S. R., Lerner, J. S., von Hippel, W., Jonides, J., & Nagda, B. A. (1998). Undergraduate student-faculty research partnerships affect student retention. *The Review of Higher Education, 22*(1), 55-72.

Hathaway, R. S., Nagda, B. A., & Gregerman, S. R. (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study. *Journal of College Student Development, 43*(5), 614.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives, 9*(2), 85-110.

Hunter, A. B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education, 91*(1), 36-74.

Ishiyama, J. (2007). Expectations and perceptions of undergraduate research mentoring: Comparing first generation, low-income white/Caucasian and African American students. *College Student Journal, 41*(3), 540–549.

Jones, M. T., Barlow, A. E., & Villarejo, M. (2010). Importance of undergraduate research for minority persistence and achievement in biology. *The Journal of Higher Education, 81*(1), 82-115.

Kuh G. D., & Love, P. G. (2000). A cultural perspective on student departure. In J.M. Braxton (Ed.), *Reworking the student departure puzzle* (pp. 196-212). Nashville: Vanderbilt University Press.

Lopatto, D. (2004). Survey of undergraduate research experiences (SURE): First findings. *Cell Biology Education, 3*(4), 270-277.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE-Life Sciences Education, 6*(4), 297-306.

National Center for Education Statistics (2011). 2004/2009 Beginning Postsecondary Students Longitudinal Study. Washington, DC: National Center for Education Statistics, US Department of Education. Accessed through NCES DataLab, https://nces.ed.gov/datalab/

Nora, A. (2004). The role of habitus and cultural capital in choosing a college, transitioning from high school to higher education, and persisting in college among minority and nonminority students. *Journal of Hispanic Higher Education, 3*(2), 180-208.

Schultz, P. W., Hernandez, P. R., Woodcock, A., Estrada, M., Chance, R. C., Aguilar, M., & Serpe, R. T. (2011). Patching the pipeline: Reducing educational disparities in the sciences through minority training programs. *Educational Evaluation and Policy Analysis, 33*(1), 95-114.

Seymour, E., Hunter, A. B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education, 88*(4), 493-534.

Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition.* Chicago: University of Chicago Press.

UW-Madison Academic Planning and Institutional Research (2016). *Retention and graduation rate tabulations (updated October 2016).* Madison, WI: Office of the Provost, University of Wisconsin-Madison. Accessed through http://apir.wisc.edu/retention.htm.

UW-Madison Academic Planning and Institutional Research (2017). *2016-2017 data digest.* Madison, WI: Office of the Provost, University of Wisconsin-Madison.

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and individual differences, 16*(1), 1-12.

What Works Clearinghouse (2014). *Procedures and standards handbook (Version 3.0).* Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf.

Zydney, A. L., Bennett, J. S., Shahid, A., & Bauer, K. W. (2002). Impact of undergraduate research experience in engineering. *Journal of Engineering Education, 91*(2), 151-157.

Appendix

Table A: Examination of compliance among those randomized to treatment, 2015-16 cohort

	Bivariate comparison			Multivariate logit regression		
	Took URS	Did not take	*p*	Coef	SE	*p*
First-generation	0.3559	0.1612	0.05	1.285	0.606	0.034
Targeted minority	0.3474	0.3548	1.00	-0.006	0.578	0.991
Female	0.5423	0.5161	0.841	0.142	0.445	0.749
ACT	28.29	28.7	0.602	0.02	0.066	0.757
Freshman	0.6271	0.5161	0.304			
In-state	0.6101	0.7096	0.404	-0.475	0.527	0.367
US citizen	0.9152	1	0.122	--	--	--
Pell recipient	0.2966	0.258	0.825	-0.508	0.638	0.426
State grant	0.2796	0.258	1.00	0.588	0.619	0.342

Note: Bivariate p computed with Fischer's exact tests (2-tailed) for dichotomous variables and t-tests (2-tailed) for continuous