



A Longitudinal Study of the Relationship Between Early Undergraduate Research and Academic Outcomes in Computer Science

Kamen Redfield*
kamenredfield@gmail.com
UC San Diego
La Jolla, CA, USA

Cynthia Lee
cbl@stanford.edu
Stanford University
Palo Alto, CA, USA

Sukham Sidhu*
ssidhu@ucsd.edu
UC San Diego
La Jolla, CA, USA

Diba Mirza
dimirza@cs.ucsb.edu
UC Santa Barbara
Santa Barbara, CA, USA

Zackary Glazewski*
zglazewski@ucsb.edu
UC Santa Barbara
Santa Barbara, CA, USA

Christine Alvarado
cjalvarado@ucsd.edu
UC San Diego
La Jolla, CA, USA

ABSTRACT

This paper reports on the longitudinal impacts of an inclusive, structured research experience program for early career undergraduates in computer science that engages a large number of students from minoritized groups. We compared academic performance and retention in the major for program participants at two large public research universities in the United States vs. a matched control group of demographically and academically similar students. We found that the retention rate of program participants was higher than the control at both universities, though not statistically significantly so. We found no significant difference in post-program GPA, and the program did not erase equity gaps in GPA by race and first generation status that existed before the program. These results help us understand the benefits and limitations of large-scale early research programs for increasing equity in computer science.

CCS CONCEPTS

• Social and professional topics → Computing education programs.

KEYWORDS

Undergraduate Research; Diversity; Early Undergraduates; Scaling Research; Computing Education

ACM Reference Format:

Kamen Redfield, Sukham Sidhu, Zackary Glazewski, Cynthia Lee, Diba Mirza, and Christine Alvarado. 2024. A Longitudinal Study of the Relationship Between Early Undergraduate Research and Academic Outcomes in Computer Science. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2024)*, March 20–23, 2024, Portland, OR, USA. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3626252.3630956>

*The first three authors contributed equally.



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

SIGCSE 2024, March 20–23, 2024, Portland, OR, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0423-9/24/03.
<https://doi.org/10.1145/3626252.3630956>

1 INTRODUCTION AND RELATED WORK

Undergraduate Research Experiences (UREs) play an important role in shaping PhD program admissions decisions, both because prior research experience is itself a high-priority qualification for admission, and also because research experience creates opportunities to have the kind of relationships with faculty mentors that lead to high-impact recommendation letters.

UREs have also been shown to promote desirable outcomes in participants' academic performance, self-efficacy, scientific identity, and retention across Science, Technology, Engineering and Math (STEM) disciplines [9–12], specifically, studies in computer science have shown benefits in retention, academic performance, and persistence to graduate school [1, 3, 5, 10, 15]. UREs may also have particular benefits for women and students from racial groups underrepresented in STEM and computer science [6, 13, 14].

Many previous studies on the benefits of research in computer science have examined traditional UREs, where individual students work closely with a faculty mentor who provides students with personalized support. These experiences are not scalable due to faculty availability to mentor UREs, and traditional UREs are typically only available to more experienced students. A key question is whether a scalable program can provide the same (or similar) benefits to a one-on-one research mentorship experience; emerging evidence suggests this is possible [3]. However, more research is needed to understand if these results will replicate.

We performed a longitudinal study of student outcomes from a scalable research program designed for early undergraduates at two large research universities (RU1 and RU2) in the United States. We compared academic outcomes several years after the program for 4 cohort years at RU1 and 3 cohort years at RU2 to a control group that was rigorously matched to be as similar to the program participants as possible. We found that program participants were retained in computing majors at higher rates, particularly women students. We did not find any notable difference in post-program GPA. We further found that there were equity gaps in GPA by first-generation status and by race that existed between program participants before the program began, and that participation in the program did not appear to change these gaps.

This work makes the following two contributions to our understanding of the benefits (and limitations) of early research experiences. First, compared to prior evaluation of scalable research programs it examines a larger group including two universities, a larger group of students, a more rigorously matched control group, and a more nuanced demographic analysis. Second, compared to other studies of undergraduate research participation which are primarily based on surveys and students’ attitudes and intentions, our study uses direct measures of students’ grades and retention in computing for several program cohorts, giving us a better understanding of the actual outcomes from research participation.

2 PROGRAM STRUCTURE

We studied the Early Research Scholars Program (ERSP) [4], a team-based research opportunity for second-year and transfer computer science and engineering students that takes place over one academic year. During the program, students attend weekly meetings where they learn basic research skills, contribute to existing CS research projects, and present their research at the end of the year. The details of this program are described in previous work [2]. Here we describe the aspects that are relevant to this work.

ERSP currently runs at eight universities across the United States. It was established in 2014 at UC San Diego (RU1) and expanded to UC Santa Barbara (RU2) in 2018 as the second implementation site. The program is designed for computing¹ students in the second year of their major, or in their first year as an incoming transfer student. The program targets students who belong to groups that are traditionally underrepresented in computing including women and non-binary students, Black, Latino/a, first-generation, and LGBTQ+ students, etc. with the goal of increasing diversity in computing.

ERSP is a large program; its size is about 10-20% of the number of students in the second year of the CS major at each university. At RU1 this equates to 50 students per year, while at RU2 it is around 30 - 40 students per year². ERSP achieves its scale through its unique group-based, dual-mentored structure. Students work in groups of four and are supported by both a technical faculty mentor who guides the direction of the research, and an additional mentor called the “central mentor” who meets weekly with all teams to help provide supplementary support including goal setting, communication, teamwork and basic research skills. In addition, students take a research methods course in the fall that is grounded in their research project to help them get up to speed quickly with the basics of their project and research in general.

3 RESEARCH QUESTIONS & APPROACH

This study sought to measure the potential impact of ERSP on the retention and academic performance for students who participated in the program at the two universities. Our research questions were:

- (1) Does ERSP participation improve retention in computing?
- (2) Does ERSP lead to improved academic outcomes?
- (3) Are there demographic differences in academic outcomes, and if so, does ERSP help mitigate these differences?

¹The program is limited to Computer Science majors at RU2, but at RU1 students in related programs (e.g. Math-CS or Cognitive Science) are also accepted.

²The first few cohorts of ERSP at RU2 included in this study were smaller than the current steady state, with 12 students in the 2018-19 cohort, 16 students in 2019-20, and 20 students in 2020-21

To address the research questions, we explored differences in grade point average (GPA) and retention between different demographic groups in ERSP. We ran some initial descriptive statistics with the ERSP cohort. Next, we chose our statistical analysis techniques. Finally, we devised our control group matching technique.

3.1 Sample

Our study included ERSP participants at two public research-focused universities in California—UC San Diego (RU1) and UC Santa Barbara (RU2)—covering cohort years cohort years 2015-16 through 2019-20 from RU1 and 2018-19 through 2020-21 from RU2. There are two reasons we used different cohort groups at each university. First, ERSP only began in 2018 at RU2. Second, the data analysis at RU1 was performed in summer 2022, while the data analysis at RU2 was performed in spring 2023, giving more time to measure academic outcomes from a later cohort at RU2. Our data set included all students in the selected cohort years who completed the program and consented to participate in the research: 190 (out of 211) at RU1 and 38 (out of 48) at RU2.

The data on ERSP participants included the following information: admit quarter; the number of units they completed and grade points for every quarter they were enrolled; binary gender; first-generation college status; and whether they were from a racial group that is underrepresented in computing (Hispanic/Latino, Native American, and/or Black/African American). At RU1, the first-generation status of three students was unknown. We also had students’ major at the time of admission to the university, when they began ERSP, and at the time of analysis. In most cases the major at the time of analysis was the major in which they had earned their degree. We had the same information for the control group except for one key difference. For the control group, we only know their major at the time of admission and the time of the study (or graduation).

3.2 Control group matching

Our goal was to measure the causal effect of ERSP on its participants. We could not use an experimental design because participants were not randomly selected to participate in the ERSP program. Instead we used a quasi-experimental design where we matched each ERSP participant to a control subject along several dimensions in order to simulate a random experiment as closely as possible.

Participants from each school were matched 1:1 without replacement with a student from a pool of students from the same university who did not participate in ERSP (the “control pool”). We decided that demographic variables were the most important to match on because the average ERSP student is much different than the average CS student, so we matched first on those and then GPA (see Table 1). At RU1, we matched the demographic variables in this order: admit term, admit department, admit type, first generation flag, race/ethnicity, and then gender. We matched on department at the time of admission and not specific major (e.g. we might match a Math-CS major with a Math major) because matching by specific major made it too difficult to obtain good matches across the other variables. At RU2 we used a slightly different order due to the smaller pool of first-generation students and the lack of transfer

Table 1: Demographic breakdown of ERSP and control group and control pool at RU1 and RU2

Inst.	RU1						RU2					
	ERSP (n=190)		Ctrl (n=190)		Overall (n=51,005)		ERSP (n=38)		Ctrl (n=38)		Overall (n=601)	
	F(%)	M(%)	F(%)	M(%)	F(%)	M(%)	F(%)	M(%)	F(%)	M(%)	F(%)	M(%)
Asian Amer.	25.79	16.32	26.32	15.79	9.09	40.93	26.32	13.16	21.05	15.70	6.66	28.95
International	20.53	6.84	20.00	7.89	4.23	17.81	7.89	5.26	13.16	10.52	4.49	18.64
Hispanic	4.74	11.58	4.74	11.05	0.95	5.57	2.63	7.89	0.00	5.26	0.83	8.12
Caucasian	4.74	5.26	4.74	5.79	1.82	15.25	10.53	15.79	2.63	26.31	3.16	16.47
African Amer.	1.58	1.05	0.00	2.11	0.22	0.82	0.00	2.63	0.00	0.00	0.00	0.99
Other	0.53	1.05	0.53	1.05	0.61	2.69	5.26	5.26	2.63	2.63	1.66	2.36
Total	57.91	42.1	56.33	43.67	16.92	83.07	52.63	47.37	39.47	60.42	17.80	82.03

Table 2: Matching similarity

Variable	RU1	RU2
First-gen	1	1
Admit term	1	1
Admit department (RU1) or admit major (RU2)	1	0.9211
Admit type (direct admit or transfer)	1	n/a
Quarter changed into computing major	-	0.9737
Ethnicity	0.979	0.8421
Gender	0.984	0.7632

students in the program. At RU2 the matching order was: first generation, admit term, admit major, quarter of entry into a computing major, ethnicity, and then gender.

Given the order of the demographic variables, the pool of ERSP students, and the control pool, our matching process proceeded as follows. For each ERSP participant, we identified potential matches from the control pool who matched exactly on each variable in the provided order. After a variable was matched, if the resulting pool of candidates was not empty, we proceeded to match on the next variable. If at any point when matching on a certain variable there were no longer any matches, then that variable was skipped and the matching continued with the next variable. After narrowing down to students that matched the demographic criteria, we selected the student from the remaining pool whose GPA was closest to the pre-program GPA of the ERSP student. For the control group, we used the GPA of the student at the same point as their matching ERSP student. For example, if the ERSP student started the program at the start of their second year, the first-year GPA of the potential control group students was considered. After matching, the ERSP student and their paired counterpart were removed from the ERSP pool and the control pool respectively.

In some cases, our matching process resulted in an exact match for all variables. The average absolute GPA difference between the participant and control groups was 0.03 for RU1 and 0.15 for RU2, with a standard deviation of for 0.22 RU1 and 0.21 for RU2. Table 2 shows the similarity between the control group and the ERSP group by variable of interest, while Table 1 shows the demographic composition of the ERSP participants, the control group, and the control pool at each university.

3.3 Dependent Variables/Outcomes of Interest

To calculate retention, we measured the proportion of ERSP students who entered the ERSP program as either computer science (CS) or computer engineering (CE) majors that either graduated with a computing major, or stayed in a computing major until the time of our study. All majors in the CS and data science departments were considered computing majors along with a few other majors outside of these departments. For RU1, the other majors that we considered computing included Math-CS, Cognitive Science with Specialization in Machine Learning and Neural Computation, Cognitive Science with Specialization in Computation, Cognitive Science with Specialization in Human Computer Interaction, and Biology with Specialization in Bioinformatics. For RU2, the additional computing majors included Statistics and Data Science.

To measure academic outcomes, we used post-program GPA as a coarse indicator of academic success. A student’s post-program GPA includes all grades after the quarter the student completed ERSP. E.g., if the student was in the 2018-19 ERSP cohort, their post-program GPA is calculated starting with the Summer 2019 courses. For the control group, we defined “post-program GPA” as the student’s GPA over the same quarters as their matched ERSP participant.

3.4 Statistical Techniques

We calculated the proportion of students retained in computing for the matched control group and ERSP group separately. We then conducted a proportions z-test for the entire group and each subgroup (broken down by gender, first-gen status, and race) and then applied the Holm method to correct for multiple statistical tests.

To analyze academic outcomes of ERSP students, we used two methods. First, to compare ERSP participant’s academic outcomes with the control, we calculated the Average Treatment Effect (ATE) on GPA. The matched control students serve as an estimate of the performance of ERSP students had they not participated in the program. Therefore, we can estimate the effect of participation by finding the difference in average performance of those who did and did not participate. We calculated the differences in post-GPA between ERSP participants and the control group and then calculated the mean ATE. To determine the significance level of the ATE we used a z-test.

We removed 6 pairs from RU1 from the analysis because one or both of the pair had either graduated or had not taken any classes after ERSP and appeared to be taking a break from school. This left us with 184 pairs in the RU1 data set. However, for students who had not taken any classes after ERSP but more than six years had passed since they entered the university, we considered them to have dropped out of the university. We kept these students in the data set and assigned them a post-program GPA of 0.

Second, we used a Mann-Whitney U test to compare the pre- and post-program GPA distributions of different demographic groups within ERSP. We divided ERSP students into groups based on first-generation status, racial under-represented group (URG) status, and gender. We compared each subgroup’s (e.g. men vs. women) GPA before and after the program using Mann Whitney U tests to test for differences. We used the Mann-Whitney U test because it is a robust, nonparametric test, and GPA distributions are heavily skewed left owing to the 4.0 scale. We then created multiple histograms to visualize the group comparisons.

4 RESULTS

4.1 RQ1: Retention

For this analysis, we focused on ERSP participants who initially entered the program as CS or CE majors at RU1 and RU2. We considered participants retained if they graduated or remained enrolled in any computing major, including those offered by other departments (see Section 3.3). This broader inclusion recognizes the strong computational focus in majors outside the CS department. If a student switched to one of these majors, we viewed it as a different perspective on the field rather than a loss of interest in CS, and thus considered them retained in computing. At RU2, all ERSP students started as CS or CE majors, except for one student who switched to CS after the program began.

At RU1, ERSP participants who initially pursued computing majors other than CS or CE were excluded from the analysis, along with their matched pairs. This decision was made because these participants were often paired with non-computing students in the control group based on their department rather than their specific major (see Section 3.2). Excluding their pairs would result in a non-equivalent control group, so our focus was specifically on CS and CE majors. It’s worth noting that both universities have limited CS and CE majors, requiring students to apply for entry, which discourages major switches.

Table 3 shows the overall computing retention rates, and retention rates by different demographic groups at both universities. Overall, we see that with one exception (first-generation college students at RU1), retention rates for ERSP students are higher than retention rates in the control group; however, retention rates for all populations are rather high and none of these differences is statistically significant after Holm adjustment. The difference in retention is largest for women students and students who identify as a member of an underrepresented racial group. Nearly all of the women who participate ERSP at both universities are retained in computing, compared to 82-87% of the women who do not participate. All but one of the participants from URGs are retained in computing across both universities, compared to a lower percent of those who did not participate. Interestingly, the retention rate for

Table 3: ERSP and control group retention rates at each institution, overall and by gender, race and first-generation status. Note that we were missing first-generation status for two students in this set at RU1.

Inst.	RU1		RU2	
	ERSP	Control	ERSP	Control
Overall	0.97 (90/93)	0.89 (83/93)	1.0 (38/38)	0.92 (35/38)
Women	0.96 (44/46)	0.82 (37/45)	1.0 (20/20)	0.87 (13/15)
Men	0.98 (46/47)	0.96 (46/48)	1.0 (18/18)	0.96 (22/23)
URG	0.93 (14/15)	0.80 (12/15)	1.0 (5/5)	0.5 (1/2)
non-URG	0.97 (76/78)	0.91 (71/78)	1.0 (33/33)	0.94 (34/36)
1st Gen	0.90 (19/21)	0.90 (19/21)	1.0 (4/4)	1.0 (4/4)
Non-1st Gen	0.99 (69/70)	0.89 (62/70)	1.0 (34/34)	0.91 (31/34)

first-generation students who do *not* participate in ERSP is (almost) identical to the retention rate for first-generation students who participate in ERSP at both universities. We are not sure if this is a general trend for first-generation students in computing at these universities or unique to this group.

RQ1 main result: Although the differences are not statistically significant (in some cases perhaps due to small N), we see similar patterns at both universities and conclude that participation in ERSP might correlate with higher retention in computing majors, particularly for students who identify as women and those from underrepresented racial groups.

4.2 RQ2: Academic Outcomes

We compared ERSP students’ post-program GPA to the control group’s post-program GPA using the Average Treatment Effect, as described in Section 3.4. The ATE of the ERSP program in our sample at both RU1 and RU2 were similar and slightly positive: 0.083 at RU1 and 0.078 at RU2, meaning that on average students participating in ERSP had higher post-program GPAs by approximately 0.08 grade points on average. However, a z-test revealed that this difference was not statistically significant ($p=0.27$ and 0.23 for RU1 and RU2, respectively).

We examined the post-program GPA along several other dimensions including separating students gender, race/ethnicity, transfer status, and first-generation college status. ATEs were small, both slightly positive and slightly negative, and not statistically significant. The same was true when we compared post-program GPA in CS courses only.

RQ2 main result: Because of the small and non-significant differences, we cannot conclude that participation in ERSP leads to higher post-program GPA.

4.3 RQ3: Demographic Differences

Finally, we examined whether pre-program academic differences existed among ERSP students in various demographic categories and explored whether ERSP participation helped mitigate these disparities. Anecdotal evidence and prior analyses have indicated equity gaps in some of our courses, with underrepresented racial groups and first-generation students receiving lower average grades. We anticipated that ERSP involvement could address these gaps by fostering a supportive peer community and enhancing students' sense of ownership in the computing field. While all participants acquire technical knowledge and self-learning skills, our hypothesis suggested that increased faculty connection and contributions to departmental research could disproportionately benefit students from underrepresented groups in computing.

Mann Whitney U tests were employed to assess pre-program GPAs of ERSP students based on first-generation status, underrepresented (racial) group (URG) status, and gender to identify potential equity gaps upon program entry. Subsequently, similar tests were conducted on post-program GPAs to determine if any gaps persisted after program completion. The analysis included only students with both pre-program and post-program GPAs. Figures 1 and 2 depict pre and post-program GPA distributions by first-generation status at RU1 and RU2, while Figures 3 and 4 show histograms by URG status. Gender-based histograms, though not included due to space constraints, revealed no apparent differences between men and women.

At both universities, we found that non-first generation students had higher pre-program GPAs (RU1 Mdn=3.68; RU2 Mdn=3.95) than first-generation students (RU1 Mdn=3.57; RU2 Mdn=3.76), though the differences were small and not (quite) statistically significant at RU1 (RU1 $p = 0.069$; RU2 $p < 0.01$). We also found that non-URG students had higher pre-program GPAs (RU1 Mdn=3.70; RU2 Mdn=3.95) than URG students (RU1 Mdn=3.43; RU2 Mdn=3.69), though this difference was only statistically significant at RU1 (RU1 $p < 0.001$; RU2 $p = 0.87$). These differences can be seen in the histograms in Figures 1–4: The distributions represented by the dark blue bars are shifted to the right of those represented by the dark red bars in all graphs. We found no significant (or apparent) difference between the pre-program GPAs of women and men.

Post-program, similar patterns persisted, evident in the light blue and pink bars in Figures 1–4. The slight differences observed in GPAs between first-generation and non-first generation, as well as between URG and non-URG, remained relatively unchanged. Notably, post-program GPA distributions closely resembled their pre-program counterparts, leading to the conclusion that ERSP participation did not mitigate GPA equity gaps as hypothesized.

RQ3 main result: Our results suggest that there may be slight GPA differences by URG-status or first-generation status prior to entry into ERSP. Participation in ERSP does not seem to exacerbate nor close these differences.

5 DISCUSSION

The literature reports many benefits of UREs, including an increased sense of belonging and interest in pursuing graduate studies. However, most studies are based on surveys and only measure students' perceived gains from participating in research. Few have specifically

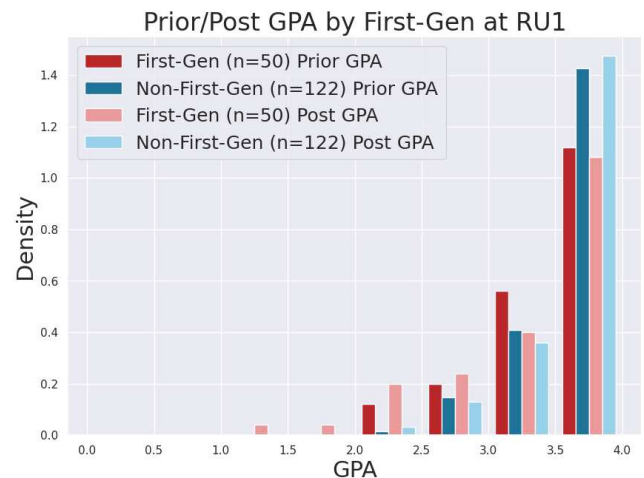


Figure 1: Distribution of pre- and post-program GPA by first generation status at RU1. Note that we did not have first-generation status for three students in the RU1 sample.

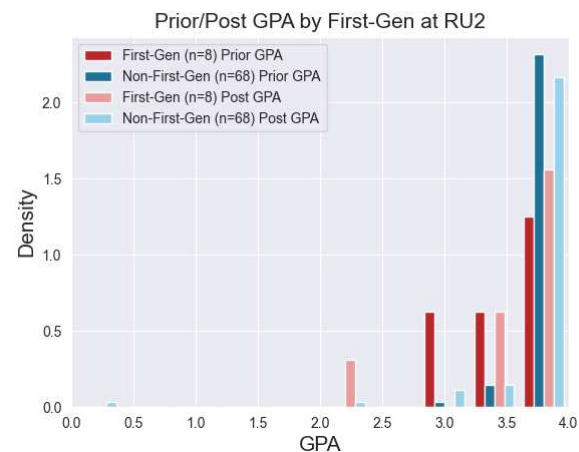


Figure 2: Distribution of pre- and post-program GPA by first generation status at RU2.

looked at the impact of UREs on hard measures i.e. retention and academic performance of STEM majors [5, 7, 8]. We build on this work and our own prior work by investigating the impact of *early research* through the *scalable* structure of ERSP on retention and academic performance (GPA) for computing students and whether these gains are different for women and students from racial groups that are underrepresented in computing.

Our findings show modest gains in retention that are not statistically significant when we use a rigorous method to control for selection bias. So, should we conclude that early research does not have much impact on retention? We do not believe this is the case. First, retention is consistently higher across both universities, particularly for women and students from underrepresented racial

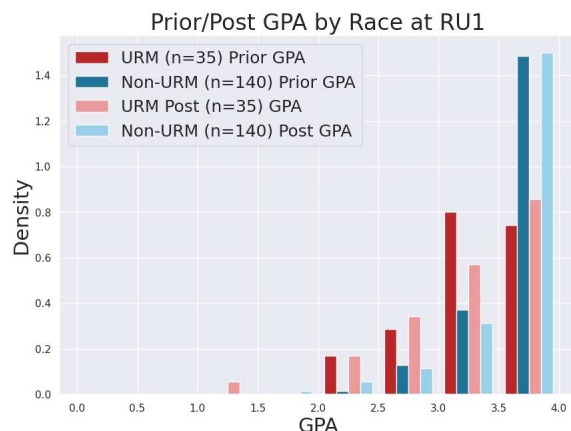


Figure 3: Distribution of pre- and post-program GPA by URG status at RU1.

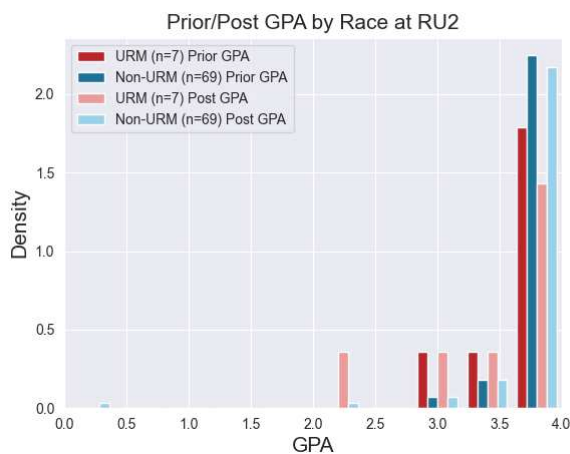


Figure 4: Distribution of pre- and post-program GPA by URG status at RU2.

groups, a result consistent with previous studies of more resource-intensive UREs. Second, our results are tempered by a ceiling effect when we control for selection bias. This is exemplified by the absolute retention numbers: 96% at RU1 and 100% at RU2. There is little room to do better. Although the matched control group has lower retention percentages (90% for RU1 and 92% for RU2), these numbers are high enough that gains are not statistically significant. It is an open question whether the ERSP program would have a larger differential effect in a context with lower baseline retention in computing, but these results are promising.

As for benefits in academic outcomes, our results are more neutral. We found no indication that participation in the program has a meaningful effect on students' grades. This result does not mean there are not other effects on students' professional success. Anecdotally we hear stories of students getting internships as a result of their ERSP experience. However, it seems that this program alone

is not enough to address structural inequities that lead to equity gaps in GPA at our universities. Finally, we are encouraged by the similar results at both universities. The fact that the impact of the program is similar at both schools shows that ERSP replicates well to a different context.

Threats to Validity

Selection bias is inherent to any program that is explicitly focused on a particular demographic subgroup, especially when students need to apply to participate. Other studies have used a control group that was constructed during the admissions process [3], but this process was not used for the program we studied due to the desire to accept all qualified applicants into the program. Although we used rigorous matching process to control for a number of important covariates, it was impossible for us to completely eliminate selection bias. Additionally, the composition of the control group was similar, but not identical, to the participant group, particularly at RU2 as the control group was drawn from a much smaller control pool.

Limitations to the repeatability of this study at other schools involve the particular characteristics and policies of RU1 and RU2. Both are large, research-intensive, highly selective public universities. Schools with other institutional and student body characteristics may have different findings. Of particular note is the ceiling effect on retention in the major. The high overall major retention is likely a consequence of the strict major admissions policies that exist at both RU1 and RU2, where students are admitted to the major in a highly competitive process, most often applying to the major directly from high school. First, this creates its own kind of selection bias in the overall pool of students to which both the ERSP students and the control group belong. Second, such high barriers to entry create strong disincentives to changing majors. Institutions with more open admissions to their computing majors may have different retention results.

6 CONCLUSION

Our work shows that structured undergraduate research programs can scale successfully, and participation in these programs may potentially have positive benefits for retention in computing. However, participation in research alone is not enough to close equity gaps in grades. Programs such as ERSP have promise, but must be part of a broader solution to change our programs to better serve all computing students.

7 ACKNOWLEDGEMENTS

We would like to thank the ERSP participants for agreeing to be part of this study. We would also like to thank the staff at UCSB and UC San Diego who helped provide the data for this study – Laurel Wilder, Associate Director, Institutional Research, Planning & Assessment, UCSB; Sara Cook, Associate Registrar, UCSB. This material is based upon work supported by the National Science Foundation under Collaborative Grant Nos. 1821521, 1821415, and 1821449.

REFERENCES

- [1] Cecilia O Alm and Reynold Bailey. 2022. Scientific Skills, Identity, and Career Aspiration Development from Early Research Experiences in Computer Science. *The Journal of Computational Science Education* 13, 1 (2022).
- [2] Christine Alvarado, Joe Hummel, Diba Mirza, Renata Revelo, and Lisa Yan. 2022. Scaling and Adapting a Program for Early Undergraduate Research in Computing. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1* (Providence, RI, USA) (SIGCSE 2022). Association for Computing Machinery, New York, NY, USA, 50–56. <https://doi.org/10.1145/3478431.3499336>
- [3] Christine Alvarado, Sergio Villazon, and Burçin Tamer. 2019. Evaluating a Scalable Program for Undergraduate CS Research. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (Toronto ON, Canada) (ICER '19). Association for Computing Machinery, New York, NY, USA, 269–277. <https://doi.org/10.1145/3291279.3339406>
- [4] Michael Barrow, Shelby Thomas, and Christine Alvarado. 2016. ERSP: A Structured CS Research Program for Early-College Students. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education* (Arequipa, Peru) (ITiCSE '16). Association for Computing Machinery, New York, NY, USA, 148–153. <https://doi.org/10.1145/2899415.2899436>
- [5] Bethany Bowling, Heather Bullen, Maureen Doyle, and John Filaseta. 2013. Retention of STEM Majors Using Early Undergraduate Research Experiences. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education* (Denver, Colorado, USA) (SIGCSE '13). Association for Computing Machinery, New York, NY, USA, 171–176. <https://doi.org/10.1145/2445196.2445249>
- [6] Mitchell J Chang, Jessica Sharkness, Sylvia Hurtado, and Christopher B Newman. 2014. What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching* 51, 5 (2014), 555–580.
- [7] John T Duncan. 2021. *Evaluating the Effects of a Biology Undergraduate Research Experience on First-Year Retention and Student Success*. Ph. D. Dissertation. Saint Leo University.
- [8] S. R. Gregerman, J. S. Lerner, W. von Hippel, J. Jonides, and B. A. Nagda. 1998. Undergraduate Student-Faculty Research Partnerships Affect Student Retention. *The Review of Higher Education* 22, 1 (1998), 55–72.
- [9] Benjamin Junge, Catherine Quinones, Jakub Kakiemek, Daniel Teodorescu, and Pat Marsteller. 2010. Promoting undergraduate interest, preparedness, and professional pursuit in the sciences: An outcomes evaluation of the SURE program at Emory University. *CBE—Life Sciences Education* 9, 2 (2010), 119–132.
- [10] Audrey Smith Rorrer, Joseph Allen, and Huifang Zuo. 2018. A National Study of Undergraduate Research Experiences in Computing: Implications for Culturally Relevant Pedagogy. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) (SIGCSE '18). Association for Computing Machinery, New York, NY, USA, 604–609. <https://doi.org/10.1145/3159450.3159510>
- [11] Susan H. Russell, Mary P. Hancock, and James McCullough. 2007. THE PIPELINE: Benefits of Undergraduate Research Experiences. *Science* 316, 5824, 548–549.
- [12] E. Seymour, A.-B. Hunter, S. L. Laursen, and T. Deantoni. 2004. Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education* 88 (2004), 493–534.
- [13] Jane G. Stout, N. Burçin Tamer, and Christine J. Alvarado. 2018. Formal Research Experiences for First Year Students: A Key to Greater Diversity in Computing?. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) (SIGCSE '18). ACM, New York, NY, USA, 693–698. <https://doi.org/10.1145/3159450.3159472>
- [14] Burçin Tamer and Jane G. Stout. 2016. Understanding How Research Experiences for Undergraduate Students May Foster Diversity in the Professorate. In *Proceedings of the 47th ACM Technical Symposium on Computer Science Education* (Memphis, Tennessee, USA) (SIGCSE '16). ACM, New York, NY, USA, 114–119. <https://doi.org/10.1145/2839509.2844573>
- [15] Rahman Tashakkori, Barry L. Kurtz, Dolores A. Parks, James B. Fenwick, and Alice A. McRae. 2011. Early Participation of CS Students in Research. In *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education* (Dallas, TX, USA) (SIGCSE '11). Association for Computing Machinery, New York, NY, USA, 63–68. <https://doi.org/10.1145/1953163.1953185>