Diversifying the Next Generation of Information Scientists: Six Years of Implementation and Outcomes for a Year-Long REU Program

Kayla M. Booth[1], Bryan Dosono[2], Elissa M. Redmiles[3], Miraida Morales[4], Michael Depew[1], Rosta Farzan[1], Everett Herman[1], Keith Trahan[1], and Cindy Tananis[1]

[1] University of Pittsburgh
kbooth,mdepew,rfarzan,everttherman,kwt2,tananis@pitt.edu
[2] Syracuse University
bdosono@syr.edu
[3] University of Maryland
eredmiles@cs.umd.edu
[4] Rutgers University
miraida.morales@rutgers.edu

Abstract. The iSchool Inclusion Institute (i3) is a Research Experience for Undergraduates (REU) program in the US designed to address underrepresentation in the information sciences. i3 is a year-long, cohort-based program that prepares undergraduate students for graduate school in information science and is rooted in a research and leadership development curriculum. Using data from six years of i3 cohorts, we present in this paper a qualitative and quantitative evaluation of the program in terms of student learning, research production, and graduate school enrollment. We find that students who participate in i3 report significant learning gains in information-science- and graduate-school-related areas and that 52% of i3 participants enroll in graduate school, over 2× the national average. Based on these and additional results, we distill recommendations for future implementations of similar programs to address underrepresentation in information science.

1 Background

Within the United States, the computer, library, and information sciences are broadly recognized as lacking racial diversity, particularly at the graduate and faculty levels [17,2,5]. This lack of diversity can stifle innovation, reduce self-efficacy and access to role models, and create social inequity. Only 5% of full professorships in science and engineering are held by underrepresented minorities; doctorates earned by minorities, though having risen slightly since 1991, remain well below 10% [9]. A non-diverse faculty and graduate student body is hardly representative of the larger population—and thus, research in the information sciences is less likely to address the most critical challenges facing minority and marginalized groups [12].

Lack of connection to a mentor has been particularly associated with lower likelihood to pursue a graduate degree in STEM or information science (IS) [1,15]. Research experiences for undergraduate programs (REUs) can help develop these connections and improve student retention, especially for minority students in computing [17,6]. A significant body of prior work has shown that REUs can increase students' interest in graduate programs and academic research careers [17,10]. Yet, minority students tend to participate in REUs at much lower rates than their White counterparts [13]—an even greater loss since prior work shows that participation in REUs more positively affects outcomes for minority students than White students [13,14].

The iSchool Inclusion Institute (i3) is an REU that specifically focuses on preparing underrepresented populations for graduate study and careers in IS and related fields. For the purposes of the program, underrepresentation refers to race, ethnicity, gender, socioeconomic status, and intersections of other identity characteristics that are not often present in the information sciences. The goal of i3 is to build a pipeline of underrepresented students to enter the professoriate and larger workforce in IS-related fields. To this end, students accepted into the program engage in hands-on, team-based IS research projects, participate in professional development seminars and instructional modules, and receive direct mentorship from faculty and graduate research advisors. Cohorts range in size from 20 to 25 students. Students from STEM and non-STEM majors are encouraged to apply to ensure greater diversity in participation from scholars interested in a broad range of academic disciplines. Our findings show that i3 has been effective: 52% of i3 students enter graduate programs, 2× the national average for all graduate programs [8], let alone IS-focused programs. As such, we distill recommendations for the deployment of similar programs at other institutions and outline suggestions for future evaluations.

2 Curriculum

The i3 curriculum is separated into three phases. In the first phase, students come to the University of Pittsburgh campus for the month-long summer Introductory Institute to learn about the information sciences, research design, programming, research topics, and graduate school. In the second phase, they participate from their home institutions in a year-long group research project in IS, conducted under the supervision of a senior doctoral student or a faculty member. Finally, a year after the Introductory Institute, students return for the third part of i3: a two-week, on-campus, Concluding Institute, during which they present their research projects and network with the incoming i3 cohort.

2.1 Introductory Institute

During the month-long Introductory Institute, students are introduced to potential research mentors, attend research talks given by CS and IS scholars, and

participate in professional development workshops. Minority scholars from universities across the country are invited to present their research, which provides a holistic overview of the topical areas within IS.

Additionally, students participate in two 10-day teaching modules focused on research design and computing methods. These modules are taught by four PhD candidates, who are selected each year as Teaching Fellows. The Fellows co-teach the modules in weeks two and three of the Introductory Institute and live alongside the students to encourage the development of mentoring relationships. Each 75-90 minute module session engages students in learning activities that provide collaborative and practical instruction. Active learning is particularly emphasized, as underrepresented students who engage in active learning opportunities such as group projects have been found to be more likely to stay in school and have better grades [13,14]. Additionally, the curriculum of the two modules is synchronized as much as possible, such that the lessons on research concepts and computational methods are complementary and enhance knowledge acquisition.

Research Design Module. This module focuses on the empirical philosophies of experimental, quasi-experimental, and non-experimental designs as a way to present the objective of research as ultimately a problem-solving endeavor [17]. Students learn how epistemology affects methodology, how to turn research problems into research questions, what makes a good research question, how to choose a research method to help answer your research question, and how to analyze data. Important topics also include how to conduct ethical research, how to develop a conceptual framework, and how to present and publish their research.

Programming Module. The programming module provides students with a toolkit of computing and quantitative analysis methods that they can draw upon during their year-long research project. Throughout the two weeks, students are exposed to and experiment with languages and tools such as Python, Tableau, and R to scrape, analyze, visualize, and present data. The programming module sessions typically involve an introductory lecture, a working session in which the students divide into two groups based on prior experience and learning pace, and a full group recap during which challenges and key points from the day are summarized.

2.2 Year-Long Research Projects

During the Introductory Institute, i3 Scholars self-select into research project teams of four to six students. Over the year, teams collaborate remotely to conduct a full IS research project with the goal of producing a conference-quality research poster or paper and an hour-long final presentation. Teams conduct a comprehensive literature review, design their study, and collect and analyze their data under the guidance of a research advisor.

2.3 Concluding Institute

After a year of working remotely, students return to i3 for the Concluding Institute, which overlaps with the last two weeks of the Introductory Institute for new students. Teams finalize and present their research papers and posters. During the Concluding Institute, the returning cohort also lives with and mentors the new cohort of students, who are preparing to embark on their year-long research projects. Finally, students also learn about the publication process and how to submit their project to research conferences in their field, such as iConference, ACM CSCW, and IEEE SeGAH.

3 Evaluation Methods

We evaluate the i3 Program using several metrics. We evaluate learning outcomes across three domains: 1) the information sciences, 2) the research process, and 3) how to navigate graduate program applications and selection.[5] Our evaluation follows a pre-post design, surveying students before and after the Introductory Institute, to enable a comparison of student understandings of IS and the research process, as well as academic plans related to IS before and after the Introductory Institute.

In addition to evaluating learning outcomes from the Introductory Institute, we evaluate the overall i3 program via two metrics: 1) the number of teams who successfully published their research projects at the end of the program, and 2) the number of program alumni currently enrolled in or graduated from graduate programs. Below we detail the data we used in our evaluation, our analysis method, and the limitations of our work.

3.1 Data and Analysis

Survey Instruments. Prior to attending the Introductory Institute, students completed the pre-institute baseline survey questionnaire which asked questions about their knowledge about IS, their interest in graduate school, and knowledge about the graduate school application process on a 5-point Likert-type scale (see Tables 1 and 2 for the questions). At the end of the Introductory Institute, students completed a post-survey questionnaire which contained the same questions, as well as open-ended questions asking them to assess the curriculum. Participants completed all survey questionnaires online via Qualtrics.

Based on the recommendations of research advisors and program alumni, the Programming and Research Design modules were added to the Introductory Institute in 2014 (year 4). Accordingly, we added open-ended questions to the post-institute survey asking Cohorts 2014-2016 for feedback on the modules. Open-ended questions allowed respondents to reflect on which aspects of the

[5] We did not collect any identifying respondent information as part of our surveys, and thus our institution's ethics review board determined that this was not human subjects research.

modules were particularly meaningful and / or helpful to them. We identified patterns and themes across responses from all three years of the research design module to complement our survey data with a rich qualitative evaluation of the students' perspective.

To analyze pre- and post-institute survey results, we use paired t-tests – a standard statistical method for comparing pre- and post-treatment sample means [11] – to evaluate whether pre- and post-responses differed significantly. This was complemented by our content analysis of the the open-ended responses. Finally, i3 maintains contact with alumni through an annual survey questionnaire, reunions and local meetups, providing letters of recommendation for applications and scholarships, and mentoring. We use data collected through alumni correspondence and via the alumni survey to calculate the number of research publications and graduate program enrollments / completions for each cohort.

3.2 Limitations

As is true of many evaluation studies, not all participants in the program chose to respond to the survey questionnaires. Thus, our data may not fully represent all student experiences. However, given that our mean response rate is 95%, we feel that our data provide a reasonable basis for evaluation. Additionally, it is possible that data validity may have suffered due to desirability bias. To mitigate desirability bias, we collected no identifying information from participants and assured them that all data would be aggregated for analysis.

4 Results

4.1 Knowledge Gains in the Information Sciences and Research

Overall, we find that participation in i3 significantly increases students' perceived knowledge about IS (see Table 1). We find a significant, nearly two out of five-point increase in perceived knowledge of subject areas in IS ($\overline{\Delta} = 1.91$) and a significant, one-point increase in perceived understanding of the impacts of technology ($\overline{\Delta} = 1.02$) following the Introductory Institute.

In addition to learning about IS fundamentals, participants reported learning about the research process. Students reported feeling that they were effectively and "thoroughly" introduced to the basic steps of conducting research, expressing sentiments such as *"I thought [the Research Design module] did a great job putting each step of the research project into perspective."*

For some students, research was new or uncomfortable territory. Participants explained how they felt after the institute, stating *"[The Teaching Fellows] aided me to understand in such small amount of time how the research process works,"* and *"I most definitely know how to plan an analysis, make a literary review, and not look stupid during a research presentation ever again."* For other participants who had prior research experience, many indicated that the program helped clarify or deepen their knowledge, explaining *"What I found helpful was the*

Question	Cohort	Baseline	Post	p-value
The diversity and breadth of subject areas within the Information Sciences.	2011	2.38	4.65	$< 0.0001*$
	2012	2.24	4.74	$< 0.0001*$
	2013	2.40	4.47	$< 0.0001*$
	2014	2.96	4.59	$< 0.0001*$
	2015	2.73	3.96	$< 0.0001*$
	2016	2.81	4.55	$< 0.0001*$
The social, political, legal and ethical impacts of technology	2011	3.29	4.50	$< 0.0001*$
	2012	3.41	4.16	$0.0201*$
	2013	3.30	4.32	$0.0016*$
	2014	3.54	4.32	$0.0008*$
	2015	3.35	4.33	$0.0004*$
	2016	3.31	4.68	$< 0.0001*$

Table 1. Mean Likert ratings (from 1-5) on knowledge of each topic. P-value statistics are calculated based on paired t-test comparison of the baseline and post survey scores; significant p-values (< 0.05) are marked with *.

clarification...[I gained] a deeper appreciation for research design. I can now clearly and confidently speak and identify various types of research design in different studies."

When commenting on the Research Design module, participants particularly highlighted two effective elements: 1) the use of concrete examples: *"clear examples and ke[pt] it simple. It made it very easy to understand the concepts and topics each day."* and 2) the direct application of concepts, *"The best takeaway from the Research Design Modules was the hands-on approach of applying the information within each session. By breaking up into groups and creating research questions, as well as mock proposals for grant money I had a better understanding of the material as well as the realization on how it gets applied."*

4.2 Interest in Pursuing Graduate Programs

In addition to assessing increases in perceived knowledge and what they valued about the new research module, we also examined students' interest in pursuing graduate programs in IS and the application process. Table 2 presents the mean across cohorts of students' perceived likelihood of pursuing a graduate degree in IS and their perceived learning about the application process.

Our results suggest that i3 students are already interested in graduate school upon their arrival to the Introductory Institute, and thus while we see an increase in interest in graduate school across all cohorts, this increase is significant for only two of six cohorts. However, we do observe a significant, nearly two-point increase in students perceived understanding of how to apply to graduate school and select which programs to which to apply ($\overline{\Delta} = 1.86$). This is especially important, as prior work has found that self-efficacy and confidence in ability are essential elements of success within IT, CS, and other STEM fields [16].

Question	Cohort	Baseline	Post	p-value
Knowledge of processes and requirements for applying to graduate school and selecting an academic program.	2011	2.62	4.75	$< 0.0001*$
	2012	3.00	4.58	$< 0.0001*$
	2013	2.45	4.72	$< 0.0001*$
	2014	2.75	4.57	$< 0.0001*$
	2015	2.88	4.42	$< 0.0001*$
	2016	2.77	4.57	$< 0.0001*$
Likelihood of pursuing a graduate degree in IS.	2011	3.90	4.40	0.0538
	2012	3.31	4.29	$0.0235*$
	2013	3.65	3.72	0.8222
	2014	3.67	4.13	0.0866
	2015	3.65	3.88	0.4132
	2016	3.65	4.29	$0.0141*$

Table 2. Mean Likert ratings (from 1-5) on knowledge of each topic. See Table 1 for detailed caption.

4.3 Research Publications and Graduate School Enrollment

To evaluate i3 in general, we analyze two non-self-reported metrics — research productivity and graduate school enrollment — which are believed to be key indicators of REU program success [3]. At the time of this papers submission, 23 of 28 research teams have successfully published and presented their projects at an academic conference, resulting in a total of 24 peer-reviewed publications (one team published twice). This level of productivity is at least comparable to and generally higher than similar REU programs in other fields [7,4]. Twenty-one of these publications were papers or posters at iConference, one was a poster at ACM Computer-Supported Cooperative Work and Social Computing (CSCW), one was a poster at IEEE Serious Games and Applications for Health (IEEE SeGAH), and one was a paper at the Serious Play Conference.

Of the 135 students who were accepted to and enrolled in Cohorts 2011-2016, 117 completed i3; 32 of these alumni are still enrolled in undergraduate programs. Of the 85 alumni who have completed an undergraduate degree, 44 (52%) are currently enrolled in or have completed graduate programs, 2× the national attendance for underrepresented minorities [8]. Of those 44 graduate students, 35 (80%) are enrolled in or have completed programs in or related to IS.

5 Discussion & Recommendations

Our analysis highlights three key findings. First, i3 scholars indicate increases in their understanding of the information sciences and self-efficacy regarding conducting research. Second, students who largely arrived with an interest in graduate school learned about programs specific to IS, and gained self-efficacy and knowledge about how to apply to and select IS graduate programs. Finally, the i3 program results in high rates of research productivity, with 23 out of 28

teams publishing a poster or paper at an academic conference, and results in high rates of graduate school enrollment: 52% of i3 students enroll in graduate programs and 80% of those graduate programs are in IS. Thus, we conclude that i3 is an effective approach to developing a pipeline for underrepresented students to enter the information sciences field.

Upon completing the program, students reported an increased awareness of the information sciences as a multi-disciplinary field. Several factors may contribute to this increase in awareness. First, students across a variety of undergraduate majors were recruited to i3. This resulted in research teams comprised of students with diverse approaches, experiences, and interests collaborating with one another to conduct research, underscoring the interdisciplinary nature of the IS field. Additionally, i3's curriculum is built on workshops delivered by numerous iSchool faculty and PhD students from different departments and schools across the country whose areas of research represent the full range of the IS field. This exposes i3 scholars to the vast topical breadth of the field and to the various methods employed. Our analysis shows this curriculum design may encourage a holistic understanding of the field in ways that a single instructor may not. The variation in research topics and methodologies helped students explore the intersections between people, information, and technology across different contexts. Participants expressed feeling they could apply what they were passionate about in their own disciplines to opportunities and areas within IS. This holistic curriculum design encourages students to see the ways in which different topics within IS and their current majors intersect.

Respondents also indicated that their skills and confidence conducting IS research increased. This perception of increased skill and ability is noteworthy in that much of the extant literature suggests that self-efficacy and confidence in their abilities is a key component of retention for underrepresented students in STEM[6]. Additionally, the majority of the research projects were accepted to conferences, which may demonstrate to students that their contributions are valuable to the field, and may enhance to their sense of belonging to a research community outside of i3. Additionally, the year-long project gave students an opportunity to work with a research advisor, which provides students with another mentor (in addition to i3 faculty and staff). This perception of skill building, successful completion of their year-long projects, connections to mentors, and subsequent publications may all contribute to increased self-efficacy, a hypothesis worthy of further investigation.

Participation in i3 also increased participants understanding of how to apply to and select graduate programs in IS. Additionally, i3 scholars come from a myriad of academic disciplines, yet the majority of scholars who are enrolled in or have completed graduate school have done so in IS. This indicates the potential of REUs to play a meaningful role in the recruitment of underrepresented students into IS.

This evaluation and analysis of the i3 program serves two functions: 1) it details a pipeline initiative that positively addresses underrepresentation within IS, providing a curriculum outline and outcome metrics that we hope will encourage

other iSchools to adopt a similar approach; and 2) it provides a basis for deeper investigation into the role of REUs and associated curriculum in IS. As these results draw from North American contexts, we encourage researchers to analyze learning outcomes of REUs in IS across the globe to extrapolate broader implementation strategies among the international iSchool community. In the next section we outline recommendations for future evaluation work.

6 Future Work

This evaluation and analysis shows that participants feel they are learning research skills and are successfully publishing their projects at peer-reviewed IS-related conferences while working with peers and mentors. i3 Scholars learn how to select and apply to graduate programs and 44 i3 alumni are enrolled in or have completed graduate school (the majority of which were IS-related programs). Though these findings suggest that the i3 program is successful in building a pipeline for underrepresented students into the information sciences, these indicators of success are only a preliminary step to fully understanding the implications of such a program.

The next step is to conduct a deeper qualitative analysis of the subjective experiences i3 Scholars attribute to helping them succeed in the program, as well as how they relate their time in the program to their post-graduate paths. Through this work, we can better understand how, why, and in what ways the different components of the program are meaningful to emerging scholars and their professional trajectories.

7 Conclusion

The results of our analysis suggest that i3 is a successful intervention for contributing to the creation of a pipeline of underrepresented students heading to IS graduate programs. It suggests that an immersive, team-based, and research-focused approach to IS is an effective way to introduce students to the breadth of the field. Our evaluation is one of the first, to our knowledge, to explore an intervention specifically designed to address underrepresentation in the information sciences and the iSchool community. Our results align with extant literature that demonstrates the potential for undergraduate research to build confidence and self-efficacy important for recruitment and retention in STEM fields. Additionally, these results provide a foundation for deeper research into how participants experience different program components, and what aspects of the program they most value and why. Further exploration of i3 offers a unique opportunity to iteratively design and evaluate a curriculum that addresses one of IS deepest challenges: inclusion and diverse representation. More broadly, future work for the iSchool community as a whole may involve the creation and evaluation of additional, similar programs at iSchools around the world as we work to "transform digital worlds" and create a more diverse community of information scientists.

References

1. C Ashcraft, E Eger, and M Friend. Girls in it: the facts. *National Center for Women & IT*, 2012.
2. William Aspray and Andrew Bernat. Recruitment and retention of underrepresented minority graduate students in computer science. In *Report on a Workshop by the Coalition to Diversity Computing*, 2000.
3. LA Beninson, J Koski, E Villa, R Faram, and SE OConnor. Evaluation of the research experiences for undergraduates (reu) sites program. *The emerging role of exosomes in stress physiology*, 2013.
4. EM Brey, MF Campanile, and NG Lederman. Evaluation of a nine year summer undergraduate research program in biomedical engineering. In *ASEE Annual Conference and Exposition*, 2015.
5. J Cuny and W Aspray. Recruitment and retention of women graduate students in computer science and engineering: Results of a workshop organized by the computing research association. *SIGCSE Bulletin*, 2002.
6. T Dahlberg, T Barnes, A Rorrer, E Powell, and L Cairco. Improving retention and graduate recruitment through immersive research experiences for undergraduates. In *SIGCSE Bulletin*. ACM, 2008.
7. DJ Follmer, SE Zappe, EW Gomez, and M Kumar. Preliminary evaluation of a research experience for undergraduates (reu) program: A methodology for examining student outcomes. In *ASEE Annual Conference and Exposition*, 2015.
8. National Center for Education Statistics. Status and trends in the education of racial and ethnic minorities.
9. National Center for Science and Engineering Statistics. Women, Minorities, and Persons with Disabilities in Science and Engineering. Technical report, 2013.
10. PR Hernandez, P Schultz, M Estrada, A Woodcock, and RC Chance. Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in stem. *Journal of educational psychology*, 2013.
11. H Hsu and PA Lachenbruch. Paired t test. *Wiley Encyclopedia of Clinical Trials*, 2008.
12. PT Jaeger and RE Franklin. The virtuous circle: Increasing diversity in lis faculties to create more inclusive library services and outreach. *Education Libraries*, 2007.
13. J Kinzie, R Gonyea, R Shoup, and G D Kuh. Promoting persistence and success of underrepresented students: Lessons for teaching and learning. Technical Report 115, 2008.
14. GD Kuh, J Kinzie, T Cruce, R Shoup, and RM Gonyea. Connecting the dots: Multi-faceted analyses of the relationships between student engagement results from the nsse, and the institutional practices and conditions that foster student success. Technical Report 547556, 2006.
15. GS May and DE Chubin. A retrospective on undergraduate engineering success for underrepresented minority students. *Journal of Engineering Education*, 2003.
16. EM Trauth, CC Cain, KD Joshi, L Kvasny, and KM Booth. The influence of gender-ethnic intersectionality on gender stereotypes about it skills and knowledge. *SIGMIS Database*, 2016.
17. J Weinman, D Jensen, and D Lopatto. Teaching computing as science in a research experience. In *SIGCSE*. ACM, 2015.