Modeling Determinants of Undergraduate Computing Students' Participation in Internships

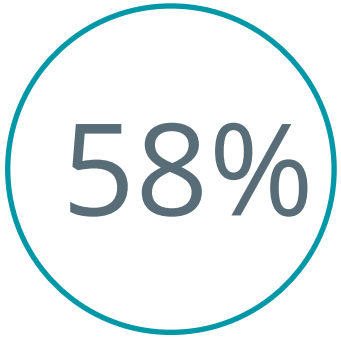
Megan Wolf, Amanpreet Kapoor, Charlie Hobson & Christina Gardner-McCune
University of Florida

Motivation

Industry Internships

Students	Employers	Computing Programs
▪ build technical skills ▪ develop professional skills ▪ gain authentic experiences ▪ explore computing pathways ▪ develop professional identities ▪ secure future employment and higher salaries	• provide cues for hiring full-time employees • offer an opportunity to evaluate potential candidates	• increase retention in computing programs • improve quality of capstone projects
(e.g., Minnes et al., 2021)	(e.g., Stepanova et al., 2021)	(e.g., Fryling et al., 2018)

Motivation



Graduating computing undergraduate students who pursue an internship in US

(Kapoor and Gardner-McCune, 2020)

Research Question

What are the factors that influence undergraduate computing students' participation in internships?

Prior Work

- Participation in internships in computing
  - importance of internships (Kapoor and Gardner-McCune, 2019)
  - student demographics (Kapoor and Gardner-McCune, 2020)
  - barriers that support/prohibit securing internships (Kapoor and Gardner-McCune, 2020)
  - used qualitative and bivariate analysis

- Modeling internships in other domains
  - race, gender, age, first-generation status, and participation in high-impact practices such as research, etc. influenced internship participation (Hoekstra, 2021)

Prior Work

- Theoretical background on identity
  - James Marcia's theory of identity development (Marcia, 1966)
  - four statuses across two dimensions: exploration and commitment
    - identity diffusion (**low** exploration, **low** commitment)
    - identity foreclosure (**low** exploration, **high** commitment)
    - identity moratorium (**high** exploration, **low** commitment)
    - identity achievement (**high** exploration, **high** commitment)

Methods

- Study Design
  - Larger study: mixed methods based on a Concurrent Triangulation Design
    - surveys and interviews
  - This work:
    - cross-sectional survey data
    - quantitative approach to analysis

- Research question
  - What are the factors that influence undergraduate computing students' participation in internships?

Methods

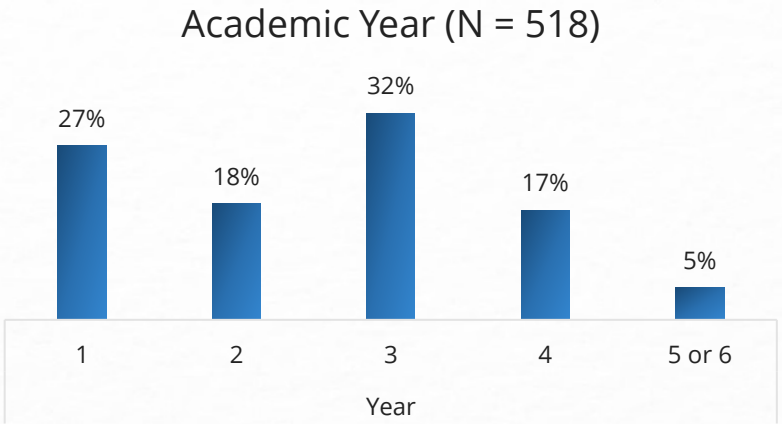
- Research Site
  - **Population:** Computing undergraduate students
  - **Sample:** students enrolled in an undergraduate computing degree program at a large public university in the US
  - **Context:**
    - admission to the site is selective
    - students can select a major when they start the program but have the flexibility to switch it at any time
    - students enrolled in CS, Computer Engineering (CE), and Digital Arts & Sciences (DAS) majors
    - participation in an internship is not mandatory before graduation

Methods

- Participants and recruitment
  - study approved by the Institutional Review Board at the research site
  - participants were recruited from CS1, CS2, software engineering, HCI, and OS courses and offered extra credit or random gift cards
  - 43% response rate (N=698, Total course enrollments=1620)
  - Discarded data:
    - students who were not in CS-related majors
    - CS minors
    - students who completed less than 80% of the survey
    - students who were not in an undergraduate program
    - students without gender classification
    - non-traditional students over age 24
    - students with a high proportion of relevant missing data

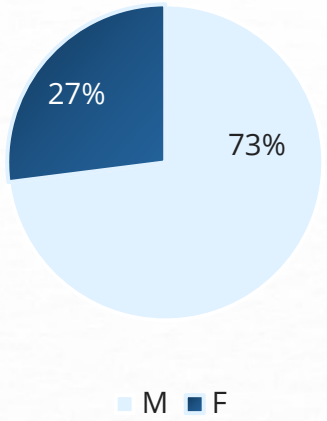
Methods

- Participants
  - 518 students enrolled in
    - CS (66%),
    - CE (26%),
    - DAS (4%), or
    - double majors including CS (4%)
  - Average age of students: 20
    - Min = 18
    - Max = 24
    - SD = 1.4

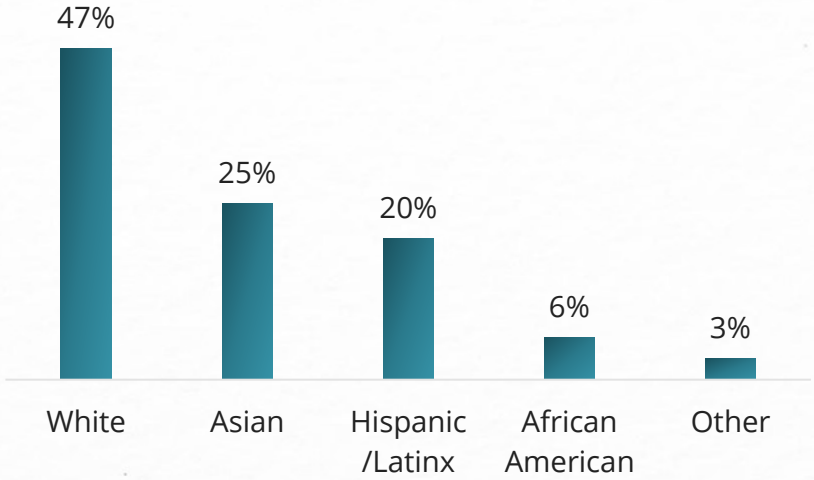


Methods

Gender (N = 518)



Race/Ethnic Identity (N = 518)



Methods

- Survey
  - 11 sections
    - demographics,
    - professional goals and identity,
    - degree program experience,
    - social support,
    - involvement in external activities
  - Questions developed from
    - findings of our prior work and pilot study
    - NCWIT Student Experience of the Major Survey
    - CRA Data Buddies Survey
    - Bennion and Adams' Extended Objective Measure of Ego Identity Status (EOM-EIS) instrument for measuring Marcia's identity statuses

Methods

- Response (dependent) variable
  - binary categorical variable
  - participation in internship(s) or co-op(s) during a student's enrollment in a degree program
  - not counting internships during high school
- Explanatory or independent variables
  - 13 independent variables

Methods

Explanatory (independent) variable descriptions

Variable Category	Independent Variable	Description (Coded value)
Demographic and Socioeconomic Factors	Household income ★	{"Less than $20,000" (1) ... "Over $150,000" (7)}
	Race/ethnicity ▲	{White/Asian (0), Underrepresented: all other ethnic and racial representations (1)}
	Gender ▲	{Male (0), Female (1)}
	Age ■	Numerical (Range: 18-24)
	Employment status ▲	{Unemployed (0), Employed - working along with the degree program (1)}
Academic Profile	GPA ■	University-level grade point average on a 4.0 scale
	High school courses in CS ▲	{No (0), Yes (1)}
	Year in school ★	{Freshman (1), Sophomore (2), Junior (3), Senior (4), Super Senior (5)}
Identity	Diffusion score ⊙	Marcia identity status composite score (scale: 6-30): Low exploration, low commitment
	Foreclosure score ⊙	Marcia identity status composite score (scale: 6-30): Low exploration, high commitment
	Moratorium score ⊙	Marcia identity status composite score (scale: 6-30): High exploration, low commitment
	Achievement score ⊙	Marcia identity status composite score (scale: 6-30): High exploration, high commitment
External Involvement	External involvement score ⊙	Composite score based on involvement in 14 activities, e.g., hackathons, clubs, etc.(scale: 0-42)

Key: Binary encoded categorical ▲ | Ordinal encoded categorical ★ | Quantitative ■ | Quantitative variable computed from ordinal scale questions ⊙

Methods

- Analysis
  - binary logistic regression model
    - participation in internship (Yes/No) is the response variable
    - 13 explanatory variables (excluded one variable later – multicollinearity)
  - The logit (i.e., the natural logarithm of an odds ratio) forms the basis of logistic regression
  - The odds ratio represents the odds that an outcome will occur (e.g., a student participates in an internship), given the presence of some factor and controlling for other predictors

Methods

- Model

$$Z = \ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{12} X_{12}$$

  - P_i is the probability of event i,
  - β_0 is the constant coefficient,
  - $X_1 \ldots X_{12}$ are the explanatory variables,
  - $\beta_1 \ldots \beta_{12}$ are coefficients of explanatory variables

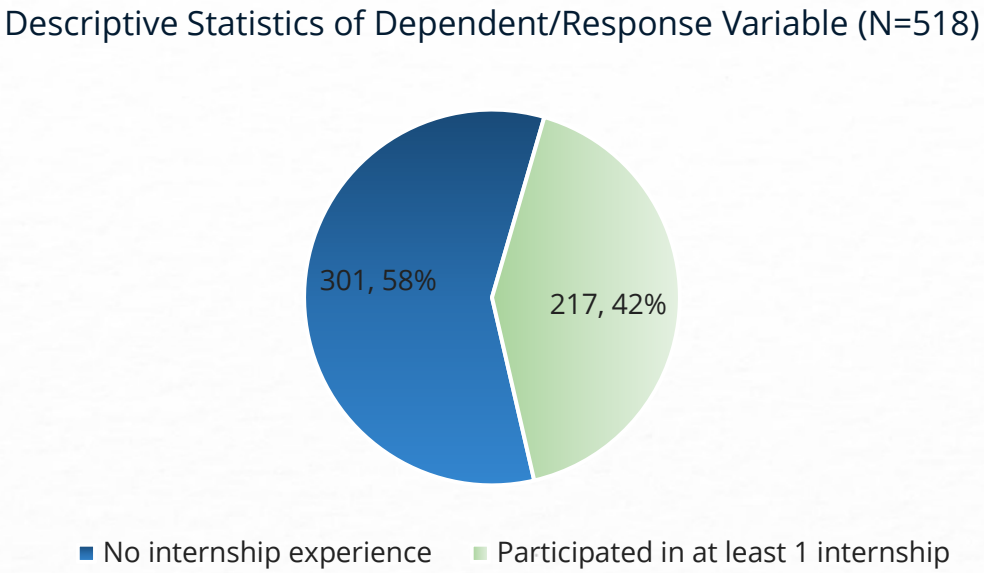
- Null and Alternate Hypothesis

$$H_0 = \beta_1 = \beta_2 = \cdots = \beta_{12} = 0$$
$$H_A = \beta_1 = \beta_2 = \cdots = \beta_{12} \neq 0$$

Methods

- Data imputation
  - handling 1.4% missing data
  - 326 of the total 23828 data points were replaced
- Multicollinearity
  - correlation coefficients such as Pearson's R (correlated if >= ±0.7)
  - variance inflation factor (correlated if VIF > 5)
  - Age was highly correlated with the year in school (Pearson's R = 0.80) and hence we excluded age

Results


Descriptive Statistics of Dependent/Response Variable (N=518)
301, 58%
217, 42%
No internship experience
Participated in at least 1 internship

Results

	Coef. β	Std. Err	Z	p > |Z|	CI for Coef. β		Odds Ratio	CI for Odds Ratio	
					[0.025	0.975]	exp(β)	5%	95%
Const	-5.53	1.60	-3.45	0.00	-8.67	-2.39	0.00	0.00	0.09
HS CS Edu.	0.29	0.21	1.36	0.18	-0.13	0.71	1.34	0.88	2.04
Employment	0.10	0.23	0.45	0.65	-0.35	0.56	1.11	0.70	1.75
Year in School	0.57	0.10	5.72	**0.00****	0.38	0.77	1.77	1.46	2.16
GPA	0.51	0.28	1.80	0.07	-0.05	1.06	1.66	0.96	2.89
Household Income	0.22	0.06	3.76	**0.00****	0.11	0.34	1.25	1.11	1.40
Gender	–0.28	0.24	-1.16	0.25	-0.76	0.19	0.75	0.47	1.22
Race	0.16	0.24	0.67	0.50	-0.31	0.63	1.18	0.73	1.88
Moratorium Score	0.01	0.03	0.48	0.64	-0.04	0.07	1.01	0.96	1.07
Diffusion Score	-0.06	0.03	-2.14	**0.03***	-0.12	-0.01	0.94	0.89	0.99
Achievement Score	0.01	0.03	0.29	0.77	-0.05	0.07	1.01	0.95	1.07
Foreclosure Score	-0.01	0.03	-0.34	0.74	-0.06	0.05	0.99	0.94	1.05
Involvement Score	0.12	0.02	6.79	**0.00****	0.09	0.16	1.13	1.09	1.17

No. of Observations: 518 | Pseudo R^2: 0.21 | LLR p-value: 4.7e-25
Df Residuals: 505 | Log-Likelihood: -279.5 | * p < 0.05
Df Model: 12 | LL-Null: -352.2 | ** p < 0.001

Results

- Model fit evaluation:
  - McFadden's pseudo-R^2 coefficient (ρ^2) = 0.21
  - indicates an excellent model fit
  - According to McFadden, "values of .2 to .4 for ρ^2 represent an excellent fit"

Discussion

- Participation in internships is associated with:
  - year in school
  - household income
  - external involvement score
  - diffusion identity score (low exploration and low commitment)
- Compared with previous work on undergraduate internship participation (Hoekstra, 2021)
  - Similar results:
    - year in school
    - external involvement
  - New results:
    - race and gender were not associated with internship participation
    - household income and diffusion identity score predict internship participation

Conclusion

- We recommend CS departments :
  - provide resources and opportunities for students' participation in activities outside the curriculum by funding student organizations, coding competitions, etc.
  - underscore the importance of involvement in external activities
  - develop support programs for students from low socio-economic backgrounds to prepare for internship recruitment process

Limitations

- EOM-EIS identity scales had lower internal consistency due to the limited number of items used for each status
- our findings could be biased by the choice of our modeling technique
- observational study and results should not be interpreted as causal relationships
- data is from a modest sample of computing undergraduates enrolled at a single institution in the US where participation in internships was optional

Questions

kapooramanpreet@ufl.edu

Thanks to the SIGCSE Special Project Grant for supporting our project ☺