The Challenges of Tracking and Understanding Student Retention in the CS Major

Chris Stephenson
Google
Mountain View, CA USA
stephensonc@google.com

Lecia Barker
Dept. of Information Science
University of Colorado Boulder, CO USA
lecia.barker@colorado.edu

Mehran Sahami
Computer Science Dept.
Stanford University Stanford, CA USA
sahami@cs.stanford.edu

Elsa Villa
College of Education
University of Texas, El Paso
El Paso, TX USA
evillae@utep.edu

Stuart Zweben
Computer Science & Engineering
The Ohio State University Columbus, OH USA
zweben.1@osu.edu

Abstract—This panel focuses on the challenges of collecting sufficient and reliable data relating to retention challenges in undergraduate computer science education programs. The panelists will explore different perspectives on the retention of women in undergraduate computing, their challenges in collecting and analyzing student data, and what they have learned from their efforts to date.

Keywords—retention, computer science education, undergraduate programs

I. INTRODUCTION

Retention of women (and students from other underrepresented groups) in higher education computer science programs remains a challenge. Despite high profile success stories at colleges like Harvey Mudd College and Carnegie Mellon University [1,2] the percentage of women at most colleges and universities remains around 18% [3]. While it appears that the pipeline is leakiest after CS1, in reality most schools have little information about the intentions of their students coming into these classes and whether or not they intend to continue.

Much retention work assumes that there is a universal set of retention issues in computing. On the contrary, students in different contexts with different backgrounds likely have vastly different reasons for staying or leaving the discipline. Truly understanding students’ intentions, and their trajectories is the first step in closing the leaky pipeline. Different institutions also experience different barriers, different students, and different sticking points.

In 2017, the ACM Education Board convened a committee of educators from different institutions of higher education to examine pipeline barriers leading to low participation of women in computer science. This committee has worked for nearly a year to gather persistence data for women and men in computer science programs in colleges and universities across the United States. In this panel, committee members will articulate the challenges to understanding students’ motivations and trajectories in a variety of contexts. These panelists will provide different viewpoints on:

- Different notions of “retention” in different contexts.
- Different types of data available for tracking and understanding students’ goals and motivations.
- Challenges of data collection specific to each institution, and for creating a unified data set to help us understand student retention across institutions.
- Insight that has been gained from initial data analysis at different institutions and across institutions.

II. PANELISTS

A. Chris Stephenson: Head of Computer Science Education Strategy, Google

Chris Stephenson works with internal Google teams and external computer science organizations globally to improve computer science teaching and learning. Stephenson has served on a number of computer science education bodies and currently co-chairs the ACM Retention Committee and is a member of the ACM Education Committee. Stephenson believes that successful interventions to current retention challenges must be data-driven and that the community will benefit significantly from increased sharing regarding successes and failures as institutions grapple with the complexities of retention data collection and analysis.

B. Lecia Barker: Associate Professor, Department of Information Science, University of Colorado and Senior Research Scientist, NCWIT

Lecia Barker will talk about the challenges in studying retention with a heterogeneous data set, the NCWIT Tracking Tool. The NCWIT Tracking Tool allows departments to enter data related to students’ entry into the major, retention in the major, and graduation, all broken down by race/ethnicity, gender, student level, and transfers. As of late 2017, about 175 departments of computer science, computer engineering, and cognates have submitted several years of data. It can be used
to study retention, but its extreme heterogeneity poses a
challenge. The data collected on different campuses also
varies a great deal. Lecia will discuss this and other sources of
variation during the panel.

C. Mehran Sahami: Professor (Teaching) Computer Science
Department, Stanford University

Mehran Sahami works at the crossroads of computer
science education, data analytics, and machine learning. He
has been involved in numerous educational projects including
co-chairing the ACM Education Board and serving as co-chair
of the ACM/IEEE-CS joint task force on Computer Science
Curricula 2013 (CS2013). Mehran has also worked on
building statistical models to aid in the analysis and
understanding of educational data. For example, he has built
models analyzing gender dynamics in CS programs as well as
created models of student performance in introductory courses
to understand population dynamics in the face of significant
enrollment increases.

D. Elsa Villa, Research Assistant Professor, College of
Education, The University of Texas at El Paso

Elsa Villa directs the Center for Education Research and
Policy Studies at UTEP, having taught at numerous levels:
grades 7 through12, community college, and university in the
disciplines of mathematics, science, education, engineering,
and computer science. Villa led a NSF-funded qualitative
study investigating identity and agency in undergraduate
Latina students. Since 1994, Villa has led and co-led
numerous STEM grants from corporate foundations and state
and federal agencies. Villa brings critical minority institution
and minority student perspectives to the panel, providing
valuable insight into the hurdles Latinas face in computer
science and engineering undergraduate studies.

E. Stuart Zweben, Professor Emeritus, The Ohio State
University

Stu Zweben coauthors the annual reports of the CRA
Taulbee Survey of doctoral-granting departments in the CS,
CE and Information areas, and the ACM NDC Study of non-
doctoral-granting departments in computing. His many years
of experience doing these surveys, his recent work with the
CRA Generation-CS Report on the current enrollment surge in
computer science, and his more than a decade in academic
administrative positions at a large public university, give him
an understanding of the type of data that realistically can be
obtained from most academic departments. Stu also has
familiarity with data about retention that have been collected
by NCWIT and ASEE, and coauthored a study of IPEDS and
other data concerning the representation of women in
academic computing programs over the past two decades.

III. PANEL STRUCTURE

Panelists from the ACM Retention Committee will report
on their efforts to collect and analyze retention data. The panel
content will be structured as follows:

- 15 minutes: Panel overview, panelist introductions,
  flow of majors at each panelist’s institution
- 20-25 minutes: Panel moderator seeds discussion
  with some or all of the following questions:
  - What are the major barriers to data
    collection at your institution?
  - What kinds of data are easy to collect and
    what are harder?
  - How do you address these challenges?
  - What retention strategies or projects have
    you started based on the results of data
    collection and analysis?
  - What novel approaches have you used to
    collect data?
- 10-15 minutes: Questions and audience discussion.

IV. CONCLUSION

Successful interventions to improve the retention of women
and underrepresented minority students in undergraduate
computer science programs require a better understanding of
the extent and the nature of the problem at many different
kinds of institutions. This panel will provide the audience with
a better understanding of the complexities of data collection
and analysis across a sampling of universities and provide
practical suggestions on the kinds of data that can and cannot
be successfully collected, how factors such as intention to major
complicate efforts to understand retention, and how data-
supported interventions can help to successfully address
retention challenges.

ACKNOWLEDGEMENTS

The panelists wish to thank ACM and the remaining
members of the ACM Retention Committee: Alison
Derbenwich Miller (co-chair), Christine Alvarado, Valerie
Barr, Tracy Camp, Erin Mindell Cannon, Carol Frieze, Colleen
Lewis, Lee Limbird, Debra Richardson, and Henry Walker.

REFERENCES

Computer Science Graduates,” retrieved September 20, 2017 from
https://www.esd.cs.cmu.edu/news/women-are-almost-half-carnegie-
mellons-incoming-computer-science-graduates

[2] O. Stanley, “Harvey Mudd College Took on Gender Bias and Now
More Than Half its Computer Science Majors are Women,” Retrieved
September 20,2017 from https://qz.com/730290/harvey-mudd-college-
took-on-gender-bias-and-now-more-than-half-its-computer-science-
majors-are-women

Continues to Produce Record Undergrad Enrollment; Graduate Degree
Production Rises at both Master’s and doctoral Level,” Computing