Retention in 2017-18 Higher Education Computing Programs in the United States

By Rodrigo Duran, *Instituto Federal de Mato Grosso do Sul*, Elizabeth K. Hawthorne, *Rider University*, Mihaela Sabin, *University of New Hampshire*, Cara Tang, *Portland Community College*, Mark Allen Weiss, *Florida International University* and Stuart H. Zweben, *The Ohio State University*

This paper builds on an earlier study of retention in U.S. computer science bachelor's programs in the 2016-17 academic year. Focusing on the 2017-18 academic year, this paper presents and analyzes comprehensive U.S. student data for the six bachelor's and three associate's computing disciplines in which ACM has produced curriculum guidelines. Disaggregating data by gender, race/ethnicity, and type of institution, the 2017-18 student retention analyses in this paper help define national benchmarks that can benefit computing educators, researchers, and academic leaders in their efforts to address diversity and inclusion issues in their institutions and provide roadmaps for further investigation.

INTRODUCTION

Retention in academic computing programs is an issue of importance to the computing community. It often arises in discussions about gender and race/ethnicity diversity in our disciplines [3], and more generally is a standard metric of program performance in terms of student participation and outcomes. In 2016, the ACM Education Board established a committee to examine retention in four-year, post-secondary computer science (CS) education programs in the United States. The committee's final report [10] stated its goals of exploring existing datasets and data challenges, identifying factors contributing to the student attrition, and recommending potential interventions to improve retention. The report noted the "paucity of data-driven analyses and recommendations in the general literature" and concluded that "empirical data to examine retention is both limited and messy." However, it suggested that data collected by the National Student Clearinghouse (NSC) Research Center [7], which the committee learned about too late to gather any of its data for analysis, could provide a much richer source of retention data for U.S. computing programs.

As a result of the committee's recommendation, and to gain more insight into the quality of NSC data, the ACM Education Board funded the acquisition of enrollment and retention data for students in 2016-17 CS bachelor's programs. The resulting study [14] covered more than 250,000 students, and analyzed enrollments and retention along the dimensions of gender, race/ethnicity, class rank, and type of institution. Satisfied with the value of the NSC data, the ACM Education Board and the ACM Committee for Computing Education in Community Colleges (CCECC) made possible the compilation of a more comprehensive NSC data set for the 2017-18 academic year, which included bachelor's and associate's degree programs in CS and other computing disciplines. Investigations focused on a particular area of computing [9] or a particular subset of institutions [12] used the 2017-18 bachelor's enrollment data. This article focuses on the retention component of the 2017-18 data in a comprehensive manner, across bachelor's and associate's programs, ACM's different computing disciplines, and a variety of subsets of institutions.

Our goal in this paper is to help the computing community understand the extent to which students are being retained in various U.S. computing programs from a national perspective.

This will help those interested in studying and improving retention to see where retention appears to be a serious problem and where it does not. The data we present can provide a benchmark for a program to assess its own retention, and to compare the results of retention improvement strategies relative to U.S. norms.

Our data do not provide any indication of why students were not retained, nor what happened to those students who were not retained. These, of course, are questions of interest. The NSC data does not have the capability to provide an answer to the former question, though it does have the capability to provide some answers to the latter question. We will discuss that further in the paper's concluding section.

NSC RETENTION DATA ACQUISITION: A CIP-BASED APPROACH

Institutions report data to NSC by student, using Classification of Instructional Program (CIP) Codes [6] to denote the student's current program of study;[1] they then use the NSC student-based data to study students who leave their institution or do not accept offers of admission. Programs get their CIP codes when they are created or when program curricular updates warrant a CIP code change. CIP codes are widely used to determine if programs are STEM programs—for instance in determining eligibility for Optional Practical Training [4].

ACM has developed bachelor's curricular guidelines for six computing disciplines: computer science (CS), computer engineering (CE), information systems (IS), information technology (IT), software engineering (SE), and cybersecurity (CY), all of which have ABET accreditation criteria [1]. ACM CCECC has published guidelines for associate's programs in CS, IT, and CY. ABET has approved of accrediting 2-year programs in CY within its Computing Accreditation Commission.

With input from computing educators with expertise in accreditation criteria and development of the ACM discipline-specific curriculum recommendations, a group of the ACM Education Advisory Committee mapped CIP codes to the six computing programs as shown in Table 1. The result was very similar to the mapping used in a prior study that used IPEDS data to examine the representation of women in computing [13].

Table 1: Mapping of CIP Codes to Computing Disciplines

DISCIPLINE	CIP CODES
CS	11.0101, 11.0701
CE	14.0901, 14.0902
IS	11.0401, 11.0501, 52.1201, 52.1206, 52.1299
IT	11.0103, 11.0201, 11.0202, 11.0301, 11.0801, 11.0802, 110804, 11.0899, 11.0901, 11.1001, 11.1002, 11.1004, 11.1005
SE	14.0903
CY	11.1003, 43.0116

[1] In this paper, CIP codes for 2010 are used, as they were the codes in effect for the period of study. The codes are updated every ten years, and some code changes will be effective in 2020.

There are, however, limitations to our approach in characterizing the representation of students within computing programs. Although the vast majority of institutions reports data to NSC, there are some institutions that do not, notably, the military academies. Furthermore, we are aware of instances where different institutions use the same CIP code for programs in different computing disciplines [9]. We have no control over such reporting discrepancies, and do not believe they materially affect our results. We hope that continuing to use a CIP-based approach to access NSC student data will make computing faculty and their departments more cognizant of the importance of the CIP codes and encourage them to take an active role in CIP code selection or revision for their computing programs.

NSC provided us with student data for each of the six bachelor's disciplines and each of the three associate's disciplines using the mapping in Table 1. The data consisted of the number of students enrolled in each discipline during the 2017-18 academic year, the number of these enrolled students who then graduated by the end of that academic year, and the number of these enrolled students who were still in the same program of study at the same institution in 2018-19. These two subsets of enrollees enabled us to study the retention of students in 2017-18 using the following definition.

> **A student enrolled in an academic computing program during a given year is considered as having been retained by the program in that year if either the student graduated from the program by the end of that year or was still in the program during the following year.**

Thus, we are in fact measuring one-year retention at the program level. We recognize that others use the term retention in different ways. For example, institutions are interested in studying retention at the institutional level. For this purpose, as long as a student stayed at the same institution, the institution considers the student as retained, even if the student has changed the program of study. NSC also affords the ability to study such institution-level characteristics, as well as multi-year retention at the program of study level. However, such studies are beyond the scope of this paper.

In addition to having data disaggregated by each computing discipline, and similar to the 2016-17 CS study, we had the data for each discipline further disaggregated by student gender and race/ethnicity, and by type of institution. Bachelor's student data also was disaggregated by class rank.

The resulting data set is by far the most comprehensive we have seen for analyzing student retention in U.S. computing programs. The data encompassed over 2,000 computing bachelor's degree programs with more than half million enrolled students, and nearly 1,200 computing associates' degree programs with more than half million enrolled students. Table 2 provides a high-level profile of the data we obtained and shows, for each degree level, the number of institutions reporting programs of study in each respective computing discipline, and the aggregate number of students enrolled

in such programs of study in 2017-18. Institutions that have programs in more than one discipline are included in each relevant column in the table; however, institutions having more than one program in a particular computing discipline (e.g., a BA and BS in CS) are counted only once per discipline in the table, with the respective program enrollments combined.

Table 2: Profile of the 2017-18 NSC Data Set.

	DISCIPLINE						
	CS	CE	IS	IT	SE	CY	TOTAL
# INSTITUTIONS REPORTING PROGRAMS OF STUDY							
Bachelor's	925	229	432	308	55	99	2,048
Associate's	381	NA	NA	657	NA	158	1,196
# STUDENTS ENROLLED							
Bachelor's	283,080	52,010	91,355	104,016	9,416	28,888	568,765
Associate's	106,356	NA	NA	136,083	NA	18,246	260,685

The next section discusses the retention data for U.S. bachelor's students; it is followed by a section discussing the retention data for U.S. associate's students. For each set of students, we analyzed retention differences by gender, race/ethnicity, and institution type. In the bachelor's section we further studied retention differences by class rank, and compared the analysis of the 2017-18 data for CS with that previously published for 2016-17 data. Each section begins with the presentation of the data and analyses aggregated across all institutions, and is then followed by presentation and analyses for different institution types. Two-tailed z-tests were performed to assess the statistical significance of retention differences throughout the paper. Following these two major sections, we provide a synthesis of major results and commentary on questions that they give rise to for further study.

RETENTION IN U.S. BACHELOR'S COMPUTING PROGRAMS

Table 3 compares retention of U.S. bachelor's students by discipline, aggregated over all institution types. The table shows both overall retention for each discipline, and breakdowns by gender, race/ethnicity, and class rank. Each cell in this table shows the percentage, of the particular category of 2017-18 enrolled students, who either graduated in 2017-18 or were still in the program in 2018-19. For some of the software engineering categories, we were unable to derive exact retention values from the data reported by NSC. These cells are noted by "NA." This is due to small cell sizes, as NSC does not report values for cells with fewer than 10 students. In all tables and figures that report race/ethnicity, we use the following abbreviations: AI for American Indian/Alaska Native, AS for Asian, BL for Black/African American, HI for Hispanic, MR for Multiple Races, NH for Native Hawaiian or other Pacific Islander, NR for Nonresident Alien, and WH for White. We use the term "Unreported" (U) to represent instances where the student's gender, race/ethnicity, or class rank was not provided, primarily because the institution chose not to report it or in some cases, because students decline to provide this information to their institution. Across the six disciplines, the range of unreported gender was 1.6–6.8% of the students with median of 4.7%, the range of unreported race/ethnicity was 15.8–37.8% with median of 22.8%, and the range of unreported class rank was 12.9–40.6% with median of 23.3%. In the analyses that follow, when we are testing gender differences between male and female students, we include students of the known gender even if their race/ethnicity is unreported. Similarly, when doing comparisons of different race/ethnicity categories, we include students of the known race/ethnicities even if their gender is unreported.

Overall retention is highest in CE and CS, and is lowest in CY and IT. Within each discipline, retention is higher for male than it is for female students. While the percentages for male and female students within each given discipline may appear similar, these gender differences are statistically significant at the 1% level for CS, IT, and CY, and significant at the 5% level for IS. The differences for CE and SE are not statistically significant. The CS result replicates that for the 2016-17 academic year [14].

Asian and Nonresident Alien students had the two highest retention rates within each discipline. Black students had the lowest retention rates in each discipline. Figure 1 illustrates that the five largest race/ethnicity categories (Asian, Black, Hispanic, White, and Nonresident Alien) follow a nearly identical retention rate ordering within each discipline. Retention differences between Black and Hispanic students are statistically significant at the 1% level except in CY, where the difference is not significant. The retention difference between

Table 3: Bachelor's Retention Percentages by Discipline – All institution types combined

	Overall	Gender			Race/Ethnicity									Class Rank				
		F	M	U	AI	AS	BL	HI	MR	NH	NR	WH	U	FR	SO	JR	SR	U
CS	75.8	74.7	75.6	80.3	69.4	83.1	64.0	71.8	73.9	72.8	79.5	76.1	75.9	61.0	73.6	80.8	83.3	74.8
CE	77.2	76.4	77.0	80.2	70.3	81.2	68.5	73.5	75.2	76.3	79.5	78.0	77.0	62.5	71.6	80.9	86.8	77.8
IS	73.6	72.9	73.7	76.9	69.7	80.3	67.7	72.5	76.3	74.3	78.8	76.0	70.1	58.4	68.8	77.1	81.5	69.9
IT	69.7	68.8	69.7	73.5	67.4	81.7	64.3	71.8	71.4	64.3	75.7	72.5	65.8	55.1	69.8	76.1	79.0	66.2
SE	74.1	72.1	74.1	79.0	NA	79.5	64.0	73.3	70.0	NA	80.9	73.3	76.1	NA	NA	NA	80.3	76.5
CY	69.3	66.9	69.8	71.7	63.6	73.1	63.4	65.8	69.3	70.1	83.3	71.1	69.0	53.6	64.2	72.9	75.9	74.6

Hispanic and White students is statistically significant at the 1% level in CS, CE, IS, and CY; there is no significant difference in IT or SE. Differences between White and Asian students are statistically significant at the 1% level except in CY, where there is no statistical significance. Differences between White and Nonresident Alien students are statistically significant at the 1% level in CS, SE and CY; they are statistically significant at the 5% level in IS and IT, and are not statistically significant in CE.

We also performed tests of the significance of retention differences by gender within the five largest race/ethnicity categories. The only significant gender difference for Asian students was in IS, where female student retention was significantly higher than that of male students at the 5% level. Among Black students, male student retention was significantly higher than that of female students at the 1% level in both IS and IT, and among Hispanic students, male student retention was significantly higher than that of female students at the 1% level in CS and SE. Among White students, female student retention was significantly higher than that of male students at the 1% level in CS and at the 5% level in IT. Finally, among Nonresident Alien students, female student retention was significantly higher than that of male students at the 5% level in CS, CE, IS and IT. These results are summarized in Table 4. Tests that were not significant are denoted by NS and tests that could not be performed due to lack of sufficient data are denoted by NA. The results illustrate that the overall higher retention of male students within each discipline is not the whole story, even when the retention difference is statistically significant. There are important subsets for whom female student retention is statistically higher than that of male students.

Except in SE, where comparisons between class ranks could not be made, retention increased by class rank within each discipline. Each of the pairwise increases between one class rank and the next was statistically significant at the 1% level.

Table 4: Retention Differences by Gender within Selected Race/Ethnicity Categories

	CS	CE	IS	IT	SE	CY
AS	NS	NS	F (5%)	NS	NS	NS
BL	NS	NS	M (1%)	M (1%)	NS	NS
HI	M (1%)	NS	NS	NS	M (1%)	NS
NR	F (5%)	F (5%)	F (5%)	F (5%)	NA	NS
WH	M (1%)	NS	NS	M (5%)	NS	NS

Tests where retention of male students was significantly higher are denoted by "M(level)" where level is 1% or 5%. Tests where retention of female students was significantly higher are denoted by "F(level)".

DISAGGREGATION BY TYPE OF INSTITUTION

The data allowed several comparisons to be made about retention at different types of institutions. These include public vs private nonprofit (Table 5),[2] Minority-Serving Institutions (MSIs)[3] vs non-MSIs (Table 6), and Carnegie R1 vs Carnegie R2 vs non-R1 or R2 institutions[4] (Figure 2) [5]. Following each table, we summarize the results of some statistical analyses performed on the data.

In CS, private nonprofits have significantly greater retention than do publics, at the 1% significance level. Within publics, retention of male students is significantly greater than that of female students, but within private nonprofits, retention of female students is significantly greater than that of male students. Each of those significance levels also is 1%. These results are consistent with those reported in [14] for CS students in 2016-17, except that in 2016-17 the retention of female students at private nonprofits, while higher than that of male students, was not significantly so.

[2] A small number of for-profit institutions also reported data to NSC, but there are too few for separate analysis. Their data is included in the overall statistics reported in this paper for both bachelor's and associate's programs.
[3] MSIs are not further subdivided by type of race/ethnicity category
[4] Carnegie R1 institutions are Doctoral-Very High Research Activity; Carnegie R2 are Doctoral–High Research Activity.

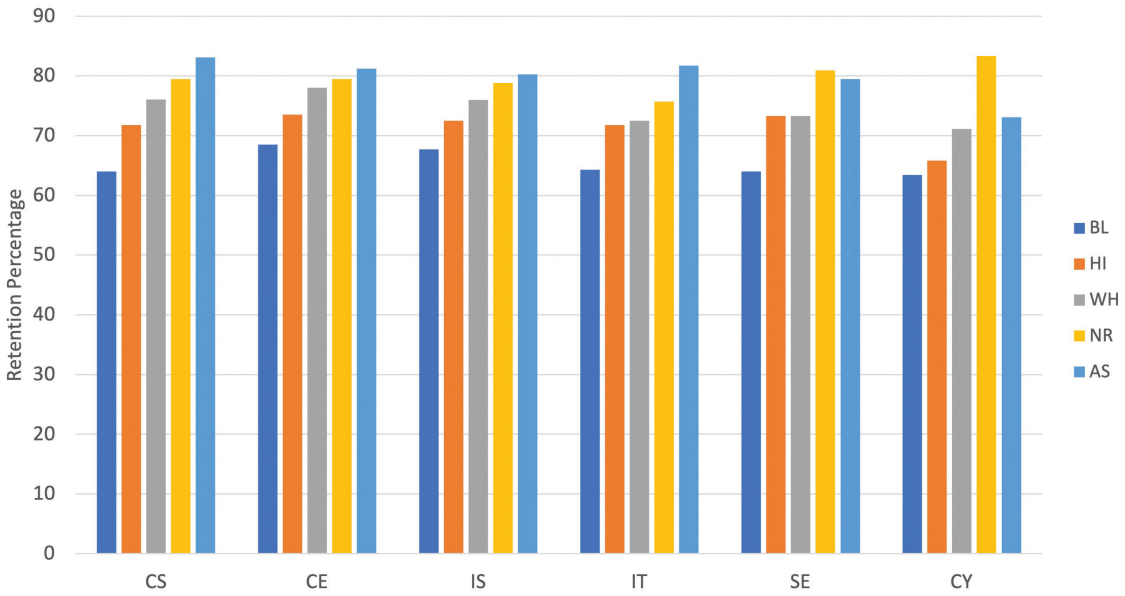

Figure 1: Similarities in Relative Retention of Selected Race/Ethnicity Categories Across Disciplines

Table 5: Bachelor's Retention Percantages by Discipline – Public vs Private Nonprofit

Public										
	#	Overall	Gender			Race/Ethnicity				
	Inst		F	M	U	AS	BL	HI	NR	WH
CS	429	76.1	74.4	76.1	79.6	83.3	64.9	72.1	79.1	76.3
CE	153	77.0	75.8	77.0	80.1	81.1	68.7	73.5	79.7	77.5
IS	251	75.6	75.8	75.3	78.1	80.9	72.6	74.5	79.1	77.4
IT	156	76.0	76.1	76.0	75.9	82.8	72.0	76.2	77.5	75.7
SE	28	72.0	69.8	71.9	79.5	78.4	62.0	71.7	82.1	70.0
CY	41	70.6	69.9	70.9	67.8	71.9	69.6	69.1	84.3	74.5

Private Nonprofit										
	#	Overall	Gender			Race/Ethnicity				
	Inst		F	M	U	AS	BL	HI	NR	WH
CS	486	76.9	77.4	76.3	83.1	83.0	66.0	72.6	81.1	76.9
CE	75	78.1	79.3	77.6	80.9	82.2	67.5	74.1	77.7	80.3
IS	170	75.4	75.4	74.9	83.8	79.9	69.9	69.5	82.1	76.8
IT	132	68.6	71.3	67.8	78.3	81.7	66.9	70.1	77.7	73.2
SE	26	81.9	80.9	82.3	76.6	85.2	72.0	80.5	NA	83.8
CY	49	74.3	70.1	75.0	82.6	77.7	63.5	66.7	NA	72.9

In CE, it also is the case that private nonprofits have significantly greater retention than do publics, at the 5% level. Similar to CS, the direction of the gender comparisons for retention at publics favors male students, while at private nonprofits it favors female students. However, neither of these differences is statistically significant.

In IS, none of these tests is significantly different.

In IT, publics have significantly greater retention than do private nonprofits, at the 1% significance level. The gender comparison for publics is not significant, while female student retention is significantly greater than that of male students at private nonprofits, at the 1% level.

Table 6: Bachelor's Retention Percentages by Discipline – MSIs vs non-MSIs

MSIs										
	#	Overall	Gender			Race/Ethnicity				
	Inst		F	M	U	AS	BL	HI	NR	WH
CS	180	77.2	74.8	77.7	77.4	83.2	66.5	73.2	77.0	77.4
CE	51	77.3	76.1	77.3	79.8	81.2	67.2	72.6	79.3	79.8
IS	83	79.7	80.0	79.6	78.8	83.1	72.5	76.5	81.3	78.8
IT	55	75.5	74.7	75.5	78.2	82.8	71.6	75.0	80.6	78.2
SE	8	75.8	69.7	76.5	82.8	85.1	NA	72.6	86.5	82.8
CY	14	76.9	75.7	77.7	70.3	73.3	65.8	73.4	NA	70.3

Non-MSIs										
	#	Overall	Gender			Race/Ethnicity				
	Inst		F	M	U	AS	BL	HI	NR	WH
CS	745	75.3	74.7	74.9	81.4	83.1	62.7	70.5	80.2	75.5
CE	178	77.1	76.5	76.9	80.4	81.2	69.3	74.7	79.5	77.3
IS	349	71.7	70.7	71.7	75.9	78.0	66.3	69.8	77.3	75.1
IT	253	68.9	68.1	69.0	72.6	81.4	63.0	70.6	74.6	72.3
SE	47	73.3	73.5	73.1	75.0	75.0	NA	73.8	78.6	73.6
CY	85	68.2	65.5	68.8	72.4	73.0	63.1	62.4	NA	69.8

In SE, private nonprofits have significantly greater retention than do publics, at the 1% level. While retention of male students at both publics and private nonprofits exceeds that of female students, the differences are not statistically significant in either case.

In CY, private nonprofits have significantly greater retention than do publics, at the 1% level. There is no gender difference for publics, but retention of male students is significantly greater than that of female students at private nonprofits, at the 1% level.

Overall retention at MSIs exceeds that at non-MSIs in each of the six disciplines. The differences are statistically significant at the 1% level in all disciplines except CE, where no statistically significant difference was present.

In CS, retention at MSIs is also significantly higher than at non-MSIs at the 1% level among male students, and among Black, Hispanic and White students. Retention of female and Asian students each show no significant difference between MSIs and non-MSIs, and retention among Nonresident Alien students is significantly higher at non-MSIs, also at the 1% level. The 2016-17 results were similar, but not identical [14]. In 2016-17, retention of Nonresident Alien students was higher at MSIs than at non-MSIs, though the difference was not statistically significant. Also, the significance level for Hispanic students was 5%, not 1%, although the difference was in favor of MSIs as it was in 2017-18.

In CE, the only significantly different retention was among White students, where MSI retention was significantly higher than non-MSI retention at the 1% level.

In IS, all comparisons are statistically significant at the 1% level in favor of MSIs with the exception of the comparison of Nonresident Alien students, which showed no significant difference.

In IT, retention at MSIs was significantly higher than at non-MSIs at the 1% level for both male and female students, and for Black, Hispanic, and White students. There was no statistically significant retention difference between MSI and non-MSIs for Asian or Nonresident Alien students.

In SE, retention of male and of Asian students each was statistically higher at MSIs at the 1% level. Retention differences for Black students could not be examined, and retention differences for female students and for Hispanic, White, and Nonresident Alien students were not significant.

Finally, in CY, retention at MSIs was significantly higher than at non-MSIs at the 1% level for both male and female students, and for Hispanic and White students. There was no statistically significant retention difference between MSIs and non-MSIs for Asian and Black students. Retention differences for Nonresident Alien students could not be examined due to lack of sufficient data.

In CS and IS, retention significantly decreased at the 1% level as the institution's research intensity decreased. The CS results are identical to those for 2016-17 [14].

In CE, retention at R1 institutions is significantly higher than that at either R2 or non-R1/R2 institutions at the 1% level, but there is no significant difference in retention between R2 and non-R1/R2 institutions.

In IT and CY, retention at R1 and R2 institutions are not significantly different, but retention at R2 institutions is significantly higher than retention at non-R1/R2 institutions at the 1% level.

For SE, retention at R1 institutions is significantly higher than that at R2 institutions at the 1% level, and retention at non-R1/R2 institutions is significantly higher than that at R1 institutions at the 1% level.

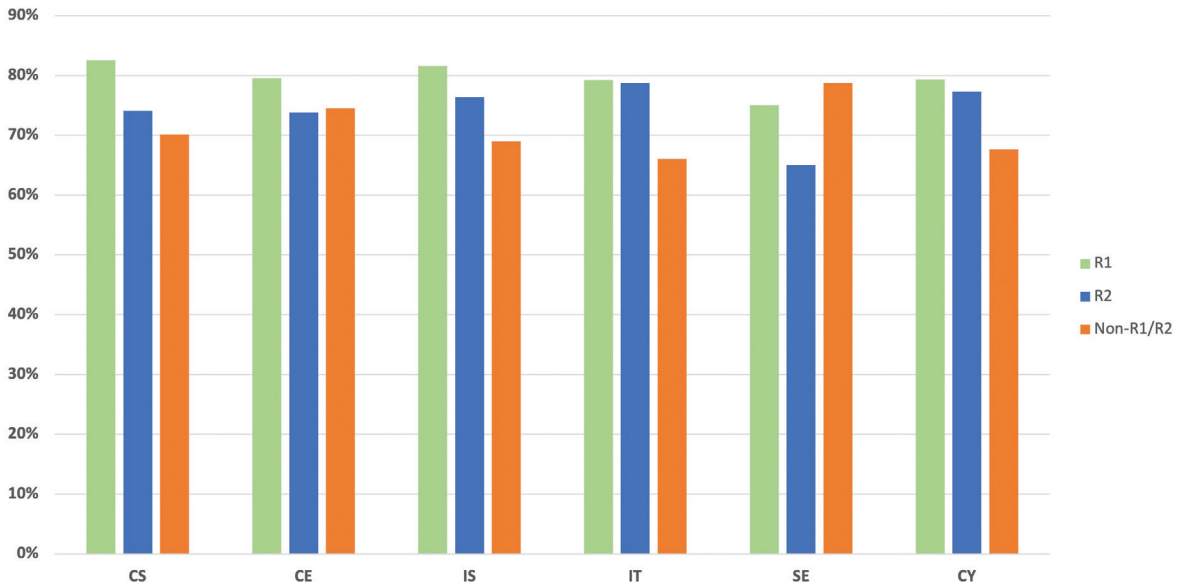

Figure 2: Bachelor's Retention by Discipline—R1 vs R2 vs non-R1/R2

RETENTION OF U.S. ASSOCIATE'S STUDENTS

Table 7 compares retention rates of associate's students by discipline, aggregated over all institution types, including breakdowns by gender and race/ethnicity. As with the Bachelor's data, each cell shows the percentage of the particular category of 2017-18 enrolled students who either graduated in 2017-18 or were still enrolled in the program in 2018-19.

Overall retention is highest in CY and lowest in CS. Within each discipline, retention is significantly higher for male students than it is for female students, at the 1% level.

In CS, Nonresident Alien and Hispanic students had the two highest retention rates, and Black students had the lowest. In IT, Nonresident Alien and White students had the highest retention rates, while Hawaiian/Pacific Islander students had the lowest. In CY, Asian and Nonresident Alien students had the highest retention rates, while Hispanic students had the lowest. Hispanic students have the most consistent retention rates across the three disciplines, rounding to 51% in all cases. In CS, this is the second highest retention rate and in CY it is the lowest. Therefore, the relative retention of associate's students in the various race/ethnicity categories lacks the uniformity across disciplines that was present in the bachelor's students. This is illustrated in Figure 3, which shows the same five race/ethnicity categories in the same order as was presented for bachelor's retention in Figure 1.

DISAGGREGATION BY TYPE OF INSTITUTIONS

Comparisons can be made about retention at different types of institutions. The preponderance of associate's level programs is at public institutions, so comparisons based on institutional control are not of particular significance. However, we can compare Minority-Serving (MSI) vs non-MSI (Table 8), as well as different Carnegie classifications of associate-degree granting institutions (Table 9). Following each table, analysis of the data is briefly discussed.

Overall retention at MSIs is higher than at non-MSIs for CS, lower for IT, and the same for CY. The differences for CS and IT are statistically significant at the 1% level.

In CS, retention at MSIs is higher than at non-MSIs among both male and female students, and among Hispanic students.

In IT, retention at MSIs is higher than at non-MSIs among unreported gender and among Nonresident Alien students.

In CY, retention at MSIs is higher than at non-MSIs among female students, and among Asian, Black, and White students.

Table 7: Retention Percentages of Associate's Students by Discipline - All Institutions

	#	#	Overall	Gender			Race/Ethnicity								
	Inst	Students		F	M	U	AI	AS	BL	HI	MR	NH	NR	WH	U
CS	381	106,356	49.3	46.1	50.0	50.2	46.5	49.7	43.8	51.0	48.5	49.3	51.9	49.5	50.0
IT	657	136,083	52.4	51.2	52.7	51.2	48.7	52.4	47.2	50.8	48.5	45.6	58.3	54.5	53.3
CY	158	18,246	55.0	52.7	55.6	56.2	53.3	60.3	53.3	51.0	53.7	52.5	58.2	57.0	52.8

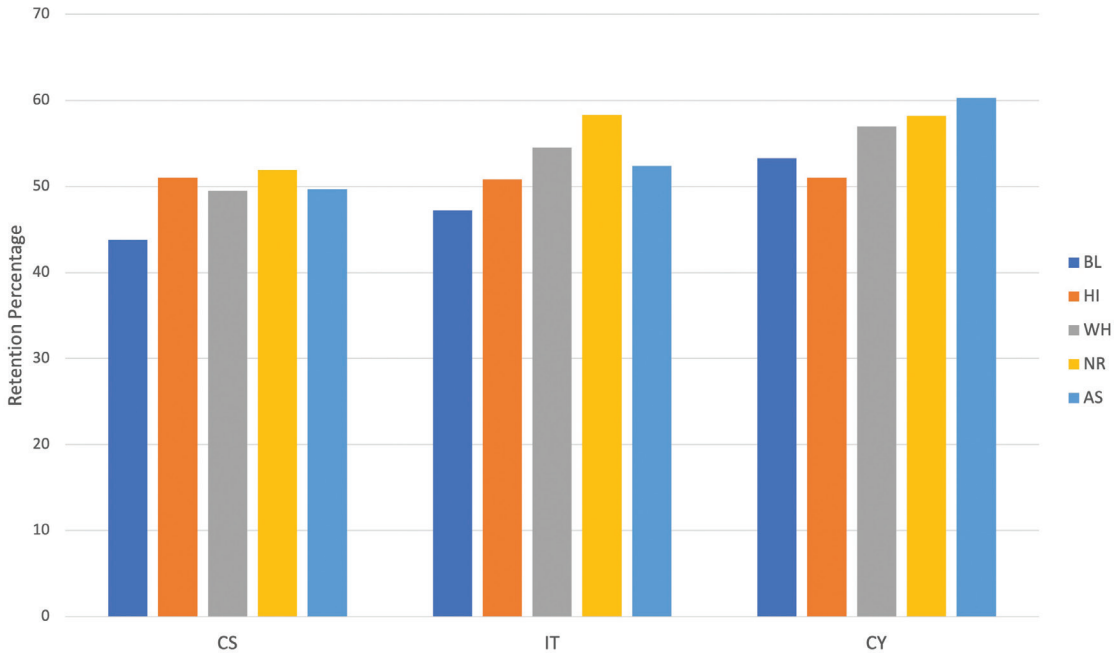

Figure 3: Relative Retention of Selected Race/Ethnicity Categories Across Disciplines in Associate's Programs.

Table 8: Retention Percentages of Associate's Students by Discipline – MSIs vs Non-MSIs

MSIs											
	#	#	Overall	Gender			Race/Ethnicity				
	Inst	Students		F	M	U	AS	BL	HI	NR	WH
CS	136	53,447	49.7	46.4	50.5	50.0	48.6	42.9	51.9	51.6	49.2
IT	176	47,973	50.5	49.3	50.6	52.9	51.5	47.2	50.2	59.4	51.2
CY	43	7,671	55.0	53.0	55.6	54.2	60.8	56.1	50.6	NA	57.5

Non-MSIs											
	#	#	Overall	Gender			Race/Ethnicity				
	Inst	Students		F	M	U	AS	BL	HI	NR	WH
CS	245	52,909	48.9	45.8	49.5	50.3	51.6	44.6	48.4	52.8	49.7
IT	481	88,110	53.4	52.3	53.8	49.9	53.9	47.3	51.6	56.0	55.3
CY	125	10,575	55.0	52.4	55.6	57.4	59.4	50.4	51.5	NA	56.8

In Table 9, the three institution types shown are defined by grouping the Carnegie categories for associate-degree granting institutions [5] as follows.

Associate-High Transfer includes:
- High Transfer-High Traditional
- High Transfer-Mixed Traditional/Nontraditional
- High Transfer-High Nontraditional

Bachelor-Associate includes:
- Mixed Baccalaureate/Associate's Colleges
- Associate's Dominant

CTE-Mixed includes:
- Mixed Transfer/Career & Technical-High Traditional
- Mixed Transfer/Career & Technical-Mixed Traditional/Nontraditional
- Mixed Transfer/Career & Technical-High Nontraditional
- High Career & Technical-High Traditional
- High Career & Technical-Mixed Traditional/Nontraditional
- High Career & Technical-High Nontraditional

Overall retention in CS is highest at Associate-High Transfer institutions (49.9%), followed by CTE-Mixed institutions (49.1%)

Table 9: Retention Percentages of Associate's Students by Discipline – Carnegie Classification (Associate-High Transfer, Bachelor-Associate, CTE-Mixed)

Associates-High Transfer											
	#	#	Overall	Gender			Race/Ethnicity				
	Inst	Students		F	M	U	AS	BL	HI	NR	WH
CS	173	64,701	49.9	47.4	50.3	52.4	50.5	44.2	51.0	51.6	49.2
IT	201	46,402	50.6	48.8	51.0	50.9	50.1	46.5	50.2	52.9	52.3
CY	58	8,919	55.3	53.6	55.4	59.9	63.4	53.8	49.9	52.6	56.3

Bachelor-Associates											
	#	#	Overall	Gender			Race/Ethnicity				
	Inst	Students		F	M	U	AS	BL	HI	NR	WH
CS	25	8,201	45.4	40.2	47.4	37.9	39.4	33.9	52.2	51.4	45.2
IT	93	657	52.3	49.9	52.9	55.6	53.3	46.6	49.6	64.2	54.7
CY	21	3,157	56.4	51.2	58.1	40.3	53.2	51.6	52.0	50.8	60.0

CTE-Mixed											
	#	#	Overall	Gender			Race/Ethnicity				
	Inst	Students		F	M	U	AS	BL	HI	NR	WH
CS	183	33,454	49.1	45.3	50.1	47.3	50.9	44.7	50.4	58.4	50.9
IT	363	61,339	53.7	53.7	53.9	49.7	55.4	48.0	52.4	56.4	55.5
CY	79	6,170	54.0	52.2	54.5	48.1	57.2	53.2	51.9	NA	56.0

and then Bachelor-Associate institutions (45.4%). Each of the pairwise comparisons (High-Transfer vs. Bach-Assoc, Bach-Assoc vs. Mixed, and High-Transfer vs. Mixed) is statistically significant at the 1% level. Retention at High-Transfer institutions also was statistically higher compared with CTE-Mixed Institutions for female students, students of unreported gender, and White students (1% level), and for Non-resident Alien students (5% level). CTE-Mixed Institutions had significantly higher retention than Bach-Assoc Institutions for each of the three gender categories, as well as for Asian, Black, and White students (all at 1% level).

Overall retention in IT is highest at CTE-Mixed institutions (53.7%), followed by Bachelor-Associate institutions (52.3%), and then Associate-High Transfer institutions (50.6%). All pairwise comparisons are significantly different at the 1% level. Retention at CTE-Mixed Institutions also was significantly higher compared with Bachelor-Associate Institutions for female students, students of unreported gender, and Hispanic students (1% level), while it was significantly lower for Non-resident Alien students (5% level). Retention at Bachelor-Associate Institutions was significantly higher compared with Associate-High Transfer institutions for male students, students of unreported gender, Non-resident Alien and White students (1% level), and Asian students (5% level).

Overall retention in CY is highest at Bachelor-Associate institutions (56.4%), followed by Associate-High Transfer institutions (55.3%), and then CTE-Mixed institutions (54.0%). Bachelor-Associate institutions are significantly different from CTE-Mixed institutions at the 5% level. The other two comparisons (Bachelors-Associates vs. Associate-High Transfer and CTE-Mixed vs. Associate-High Transfer) are not significant. Retention at Bachelor-Associate institutions is significantly higher compared with Associate-High Transfer institutions for male students, and for Asian and White students (each at the 5% level), while it is significantly lower for students of unreported gender (1% level). Retention at Associate-High Transfer institutions was significantly higher compared with CTE Mixed institutions only for Asian students (5% level). Bachelor-Associate institutions had significantly higher retention compared with CTE-Mixed institutions for male students (1% level) and White students (5% level), and significantly lower retention for Non-resident Alien students (1% level).

SYNTHESIS AND USE OF KEY FINDINGS

The previous sections delineated the detailed results of our analysis of the data. In this section, we focus on the high points of what the data revealed, and offer some suggestions for how this information can be of use to our community. As mentioned earlier, the purpose of our study was not to offer explanations of why the data show what they do, but rather to provide concrete evidence of the nature of retention in both bachelor's and associate's computing programs—evidence that can be used by other researchers, academic leaders, industry, and others to inform and ground their work.

To what extent does the data show that retention in computing programs really is a problem? We can explore this question from a variety of perspectives.

… the purpose of our study was not to offer explanations of why the data show what they do, but rather to provide concrete evidence of the nature of retention in both bachelor's and associate's computing programs—evidence that can be used by other researchers, academic leaders, industry, and others to inform and ground their work.

Our various analyses included tests of statistical significance of retention differences. Because the sizes of most of the populations studied are very large, what may appear to be small absolute differences in retention percentages in fact end up being significant. If one's goal is to reach a situation where there is no statistical difference in retention between two populations, as might be desired across gender or race/ethnicity categories, these significance tests inform decisions concerning which situations need attention, while a simple "eyeballing" of actual retention values may reach different conclusions. This is not to say that statistically equalizing retention across genders or race/ethnicity categories will itself have a major impact on diversity in computing. It is possible to have statistically equalized retention and also have serious imbalances in the representation across the gender and race/ethnicity categories relative to their population in society. But exacerbating imbalances in representation with differential retention that further weakens this representation compounds the diversity problem.

From the perspective of gender differences, the data affirmed that there are many significant differences in gender retention, generally in favor of retaining male students. While we were not surprised by this, given previous research involving multiple institutions (see, e.g., [2] and [11]), we were able to quantify these differences on a national scale as well as showing where the gender retention differences are not significant. Furthermore, by looking more deeply at institutional characteristics, racial/ethnicity considerations, and different computing disciplines, we showed that there are gender retention differences that favor female students. For example, retention of non-resident female students is actually higher than that of non-resident

male students in many of the disciplines. At private non-profit institutions, retention of female students aggregated across all race/ethnicity categories is higher than that of male students in CS and IT, the two disciplines with the largest population of enrolled students. But this is not true at public institutions. The reasons for these results are worthy of exploration and understanding. Other research has identified factors that appear to influence retention of female CS students [8]; do these factors relate to the variables in our study and how do they apply to disciplines other than CS? So, is retention of female students a greater problem than that of male students? The answer appears to be "it depends."

Table 10: Representation of Black Students Among Freshmen by Discipline in 2017-18 Bachelor's Programs

	Percentage of Freshman for whom Race/ Ethnicity is Known
CS	13.4%
CE	8.5%
IS	23.8%
IT	18.1%
SE	6.1%
CY	18.0%

From a race/ethnicity perspective the data showed that, in bachelor's programs, Black students had lower retention than the other major race/ethnicity categories, no matter which discipline was considered. While this may support a conclusion that retention of Black students is a significant problem in lack of diversity in computing, there is more that can be learned from NSC data about how one might approach the problem. For example, at some institutions, Black students are underrepresented in the computing program(s) from day one. We examined the NSC data by class rank and saw that this typically is the case in the engineering areas of computing (i.e., CE and SE programs), where representation among the freshman class is less than 10% (see Table 10). Approaches to increasing diversity may be different for programs in this category than they are in programs where there is a reasonable representation of Black students in the freshman class. In the IT discipline, for example, Black students comprised 16.7% of the total enrollment in 2017-18 aggregated across all class ranks. This was above the 13.8% national average of Black student representation among all undergraduate students in all disciplines for whom race/ethnicity was known. In the freshman and sophomore class ranks, the representation was above this national average, while at the junior and senior class ranks, the representation was below the national average (18.1% for freshman, decreasing steadily to 12.3% for seniors) [9]. This suggests that, on a national scale, Black students may not be underrepresented in the introductory IT classes. For Black students in IT, it could be more about keeping students in the computing program than it is getting them to select the program in the first place, while for Black students in CE, there is a serious issue relative to their selection of this discipline. On the other hand, it is possible that the set of institutions that offer an IT degree has overall Black student representation higher than the national average of 13.8%, which could lead to the conclusion that both representation and retention are the issue for IT.

From the perspective of gender differences, the data affirmed that there are many significant differences in gender retention, generally in favor of retaining male students. While we were not surprised by this, given previous research involving multiple institutions … we were able to quantify these differences on a national scale as well as showing where the gender retention differences are not significant.

MSIs had significantly higher retention of bachelor's students than did non-MSIs in almost every discipline. However, associate's IT programs had significantly higher retention at non-MSIs. With respect to the comparative retention among the major race/ethnicity categories, there was more variability in the retention results across the three associate's level disciplines than there was across the six bachelor's disciplines. It would be useful to understand this as well.

CONCLUSION

We have provided a national-level perspective on retention in higher education computing programs in the United States, based on the very comprehensive data collected by the National Student Clearinghouse (NSC) Research Center about students enrolled in such programs during 2017-18. Our study reaffirms the value of the NSC as a rich source of data for the computing community.

Our analyses illustrate retention differences by gender, race/ethnicity, and type of institutions. The analyses also illustrate how retention differs across the various computing disciplines, and how retention differs between bachelor's and associate's programs. The results reported here can serve as a benchmark for a computing program to compare, on a national scale, its retention with retention in programs in the same discipline

at institutions of similar types. By providing national level retention benchmarks relative to gender and race/ethnicity, the results also can help to ascertain the effectiveness of programs that strive to improve diversity and inclusion in computing. As seen in the previous section, the retention differences we observed pose several interesting research questions to help us understand why these differences may be present.

Because we provided only one year's worth of retention data in this study, our results, even when statistically significant, should not be generalized without further analysis. Many, but not all, of the U.S. CS bachelor's results were consistent with those from the previous year's data. Going forward, we expect to provide reports from which trends can be seen. The existence of comprehensive data reported herein should be particularly interesting to compare with data during the COVID-19 period that began in academic year 2019-20. We are analyzing 2018-19 data in order to have two years of pre-COVID-19 retention data as a baseline for such a comparison. Once COVID-19 era data is available, we expect to provide a comparison of pre-COVID-19 and COVID-19 era data.

Other studies of interest include an investigation of students who were not retained. NSC does not have information about why a student was not retained. But since it is capable of tracking students through their enrollments in various academic programs, we can learn whether an unretained student went to another program at their institution or whether the student changed institutions. The student's new program of study is also of interest; perhaps the student went to another computing program at the same institution, or one in another STEM area. Perhaps the student went to another institution to study in the same discipline. Yet another question often asked concerns how many students who graduated actually went to higher education programs at the next level (i.e., associate's graduates going to bachelor's programs and bachelor's graduates going to graduate programs). The ACM Education Board has funded the acquisition of such data from NSC for the same students enrolled in 2017-18. We are preparing a report on the results of studying this data.

We recognize that our study is limited to computing programs in the United States. We also have interest in comparable data from other parts of the world, and hope that this study motivates the reporting of similar data analyses from other countries so that the computing community may learn about computing enrollment and retention patterns globally. ❖

References

1. ABET, Inc. 2020. July 31; https:// abet.org. Accessed 2021 Apr 15.
2. Cohoon, J.M. Toward Improving Female Retention in the Computer Science Major. *Communications of the ACM*, May (2001): 108–114.
3. Computing Research Association. *Generation CS: CS Undergraduate Enrollments Surge Since 2006*. (Washington, DC: Computing Research Association, 2017).
4. Department of Homeland Security. *Study in the States*, 2021; https://studyinthestates.dhs.gov/stem-opt-hub/eligible-cip-codes-for-the-stem-opt-extension. Accessed 2021 April 15.
5. Indiana University Center for Postsecondary Research. 2018. *The Carnegie Classification of Institutions of Higher Education*; http://carnegieclassifications.iu.edu. Accessed 2020 Aug 20.
6. National Center for Education Statistics. 2020. *IPEDS Classification of Instructional Programs*. July 31. https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=55. Accessed 2021 Apr 15.
7. National Student Clearinghouse Research Center; https://nscresearchcenter.org. Accessed 2021 Apr 15.
8. Pantik, K. and Clarke-Midura, J. Factors that influence retention of women in the computer science major: A systematic literature review. *Journal of Women and Minorities in Science and Engineering*, 2019: 119–145.
9. Sabin,M., Zweben, S., Lunt, B., and Raj, R. Evaluating Student Participation in Information Technology Programs in the U.S. *SIGITE'20: Proceedings of the 21st Annual Conference on Information Technology Education*. Virtual Event: ACM (2020): 93–99.
10. Stephenson, C., Derbenwick Miller, A., Alvarado, C., Barker, L., Barr, V., Camp, T., Frieze, C., Lewis, C., Cannon Mindell, E., Limbird, L., Richardson, D., Sahami, M., Villa, E., Walker, H., and Zweben, S. *Retention in Computer Science Undergraduate Programs in the U.S.: Data Challenges and Promising Interventions*. (New York, NY, ACM, 2018).
11. Tamir, N.B. and Stout, J.G. Recruitment and Retention of Undergraduate Students in Computing: Patterns by Gender and Race/ Ethnicity (2016); http://cra.org/cerp/research-findings/. Accessed 2021 Jan 20.
12. Zweben, S., Tims, J., and Timanovsky, Y. 2020. "ACM-NDC Study 2019-20: Eighth Annual Survey of Non-Doctoral-Granting Departments in Computing." *ACM Inroads*, September (2020): 26–37.
13. Zweben, S. and Bizot, B. Representation of Women in Postsecondary Computing: Disciplinary, Institutional and Individual Characteristics. *Computing in Science & Engineering*, March/April (2016): 40–56.
14. Zweben, S. Enrollment and Retention in U.S. Computer Science Bachelor's Programs in 2016-17. *ACM Inroads*, December (2019): 47–59.

Rodrigo Duran
Instituto Federal de Mato Grosso do Sul, Brazil
rodrigo.duran@ifms.edu.br

Elizabeth K. Hawthorne
Rider University
Department of Computer Science and Physics
2083 Lawrenceville Road, Lawrenceville, NJ 08648
ehawthorne@rider.edu

Mihaela Sabin
University of New Hampshire
88 Commercial Street, Manchester, NH 03101
mihaela.sabin@unh.edu

Cara Tang
Portland Community College
12000 SW 49th Ave.
Sylvania, TCB 312, Portland, OR 97219
cara.tang@pcc.edu

Mark Allen Weiss
Florida International University
College of Engineering and Computing
10555 West Flagler Street, Miami, FL 33174
weiss@fiu.edu

Stuart H. Zweben
Department of Computer Science and Engineering
The Ohio State University
2015 Neil Avenue, Columbus, OH 43210
zweben.1@osu.edu

DOI: 10.1145/3448356