

Findings from a Pipeline Study based on Graduate Admissions Records

Susanne Hambrusch, Ran Libeskind-Hadas, Amy Csizmar Dalal, Michael Ernst,
Valerie Taylor, Charles Isbell, Elijah Cameron

July 22, 2014 - **DRAFT**

1. Introduction

The CRA-E is tasked with supporting CRA's efforts to maintain a healthy research pipeline in computing fields. To that end, the CRA-E has been studying the baccalaureate origins of domestic students who pursue Ph.D.'s in computer science.

The CRA-E's first pipeline study was based on publicly available data, notably from NSF's WebCASPAR database [1]. The study was published in the CRA's newsletter, Computing Research News, in 2013 [2]. Among the findings of the 2013 report was that in 2010, over 70% of Ph.D.'s awarded to domestic students went to students who completed their undergraduate studies at research universities, 15% at master's institutions, 11% at four-year colleges, and about 4% from other/unknown institutions. In addition, a small number of baccalaureate institutions produce a disproportionately large number of undergraduates who go on to get Ph.D.'s in computer science: Approximately 50% of domestic students come from 54 schools of baccalaureate origin and the other 50% come from over 747 institutions. These percentages have not changed significantly since 2000.

The current study seeks to obtain a clearer picture of the baccalaureate origins of domestic Ph.D. students applying to graduate programs by using graduate admissions records from a representative set of graduate schools. This study was begun by Charles Isbell and Elijah Cameron at Georgia Tech. Data from five institutions in the original data set are used in this study which was extended by the authors of this report to 14 institutions.

Fourteen computer science departments provided a total of 7,032 graduate admissions records for students applying to their Ph.D. program between 2007 and 2013. These schools ranged in rank from approximately 5 to 70 using the 2014 U.S. News and World Report (USNWR) ranking of graduate programs in computer science [3]. All 14 institutions providing application records have the Carnegie Classification RU/VH (Research Universities/Very High research activity) [4].

While the results of the current study are generally consistent with the findings in the 2013 report, some additional findings emerge:

- Undergraduates from the highest ranked research universities and selective liberal arts colleges have significantly higher admission rates than students from other institutions.

- Top 10 graduate programs have uniformly high matriculation rates for admitted students, except for students from selective liberal arts colleges. Graduate programs ranked 11+ show noticeable difficulty matriculating students from the selective universities and colleges.
- The average profile (with respect to GPA, GRE, admission rates, and matriculation rates) of females is very close to that of all students. For African-American and Hispanic/Latino students, the profile of applicants differs from that of the entire applicant pool, but the profile of admitted students from these groups is quite similar to that of all admitted students.

We expect the results of this study to be useful both to graduate programs in examining their own admissions and recruiting practices and to undergraduate programs in advising their students on admissions profiles of graduate schools.

2. Methodology

Throughout this report, we partition the 14 participating institutions into those that are ranked 1-10 and those that are ranked 11+. We henceforth refer to these institutions as the “consumers” since they take students into their graduate programs. The decision to partition the consumers into the two groups, 1-10 and 11+, was made based on observed admission characteristics and splitting into a roughly equal number of applications in each of the two groups. The institutions where students complete their undergraduate studies are henceforth called the “producers”.

We recognize that these two groupings are imperfect and subject to debate, but they do provide a simple means for partitioning the data. (The top few most selective institutions were not surveyed since their profiles can be expected to look significantly different from those of almost all other institutions.) The total number of applications (or “records”) in this study is shown in Table 2.1 below.

Applications to four departments ranked 1-10	3670	52%
Applications to ten departments ranked 11+	3362	48%
TOTAL	7032	100%

Table 2.1: Number of application records in this study by consumer group (1-10 and 11+).

The admissions records come from students having completed a baccalaureate degree at a producer department and each record includes the following:

- The year that the student applied
- The name of the consumer institution (the school to which the student is applying)
- The name of the producer institution (the schools from which the applicant received the undergraduate degree)
- The ethnicity and gender of the applicant (if reported)

- GPA and GRE scores (if reported)
- The decision of whether or not the student was admitted and, if admitted, whether or not the student matriculated to that graduate program.

While the dataset used in this study is quite large, it also has some inherent limitations. For example, students who applied to multiple doctoral programs in our study will result in repeated records. Since the records do not have names associated with them, it is not possible to reliably detect or avoid these duplicates. In addition, these records do not include important inputs in the admission process such as letters of recommendation, statements of purpose, and research achievements.

Our analysis partitions the producers into seven categories using the U.S. News and World Reports (US NWR) rankings of graduate schools in computer science, the U.S. News ranking of liberal arts colleges, and the Carnegie classifications of institutions. The Carnegie classifications are widely-used categories of U.S. colleges and universities. Using the 2005 Basic Classification of the Carnegie Foundation together with the U.S. News and World Reports rankings, we define seven categories of producers in Table 2.2. We note that the “TOP4” was made into a separate category because the 2013 WebCASPAR study shows that these four schools alone are the producers of approximately 10% of all domestic Ph.D. students.

Category Name	# of Schools	Explanation
TOP4	4	Top 4 CS departments (CMU, MIT, Stanford, UC Berkeley)
TOP25-TOP4	21	21 CS Departments ranked in the top 25 by US NWR and excluding (“minus”) the Top 4 CS departments
RU/VH-TOP25	83	Universities with a Carnegie classification of “Very High” research (RU/VH Carnegie classification) excluding (“minus”) the Top 25 CS Departments in the US NWR ranking
RU/H	99	Universities whose Carnegie classification is “High” research (RU/H)
Master’s	724	Schools with a Carnegie classification of master’s institution
TOP 25 LA+	27	Schools ranked in the Top 25 Liberal Arts Colleges by US NWR plus Olin and Rose-Hulman
BAC/A&S-TOP25LA	246	Schools with a Carnegie classification of Baccalaureate Arts and Sciences four-year college excluding (“minus”) the Top 25 Liberal Arts Colleges

Table 2.2: The seven producer groups providing 91.6% of the applications. The number of institutions is shown in the second column.

The applications from the seven baccalaureate producer groups correspond to 6,440 of the 7,032 application records. We chose the groups based on how we expect admission committees to assess the baccalaureate institutions of their applicants. Application from CMU, MIT, Stanford, and UC Berkeley are generally sought after as are applications from top Liberal Arts institutions. We split the remaining RU/VH departments into the group of top 25 (minus the top 4) and the rest to avoid having one group with over half of the applications. We keep RU/H and master’s institutions as two groups and place all non-top

25 liberal arts institutions into one group. This partitioning excludes 8.4% of the application records. They correspond to applications from students receiving their baccalaureate degree from an international institution or from US institution with a Carnegie classification having very few applications (includes DRU institutions).

Throughout, *admission rate* refers to the percentage of the students admitted and *matriculation rate* refers to the percentage of admitted students matriculating (i.e., accepting admission, also known as “yield on admission”). In addition, percentages with respect to the application pool are given as relevant.

3. Applied, Admitted, and Matriculated

Table 3.1 provides an overview of the percentage of students admitted and matriculating for all 14 departments in our study, divided into the two consumer groups. The overall admission rate is 35% and, not surprisingly, it differs between the two consumer groups. The admission rate for departments ranked 1-10 has an average of 25% and it ranges for the departments in that group from 20% to 33%. The admission rate for departments ranked 11+ has an average of 46% and it ranges from 29% to 70%.

	applications	admission rate	matriculation rate
all departments	7032	35%	40%
departments 1-10	3670	25%	47%
departments 11+	3362	46%	36%

Table 3.1: Admission and matriculation rates for the two consumer groups.

Table 3.2 shows the number of applications coming from each of the seven producer groups, along with the number of students admitted and the admission rate.

Producer Group	Applied	Admitted	Admission Rate
TOP 4	620	304	49%
TOP 25 RU - TOP 4	1637	642	39%
RU/VH - TOP 25	1835	654	36%
RU/H	667	304	46%
Master’s	807	223	28%
TOP 25 LA+	421	185	44%
BAC/A&S-TOP25LA	453	144	32%
TOTAL	6440	2456	38%

Table 3.2: Application and admission numbers and admission rates for all 14 departments by the seven producer groups.

The first pipeline study, using WebCASPAR data, found that about 70% of domestic Ph.D. students completed their undergraduate studies at research universities (RU/VH and RU/H), 15% at master’s institutions, 11% at four-year colleges, and about 4% from the other/unknown category [1]. The study did not provide a finer analysis into producer groups. As shown in Table 3.2, this distribution is similar for the dataset used in the current study: 74% of all applications come from research institutions, 13% from master’s institutions and 14% from four-year Liberal Arts colleges. Looking at the distribution of application records receiving admission offers we find that 78% come from RU/VH and RU/H institutions, 9% for master’s, and 14% from 4-year Liberal Arts institutions.

Figure 3.1 provides a summary of the admission rates from each of the seven baccalaureate producer groups (x-axis) to all participating consumers (graduate schools) ranked 1-10 and those ranked 11+. While it is not surprising that schools ranked 1-10 have lower admission rates than schools ranked 11+, the disparity in these rates is particularly notable for some groups such as Carnegie classified RU/H departments where the 1-10 admission rates are 9% and the 11+ rates are 42%.

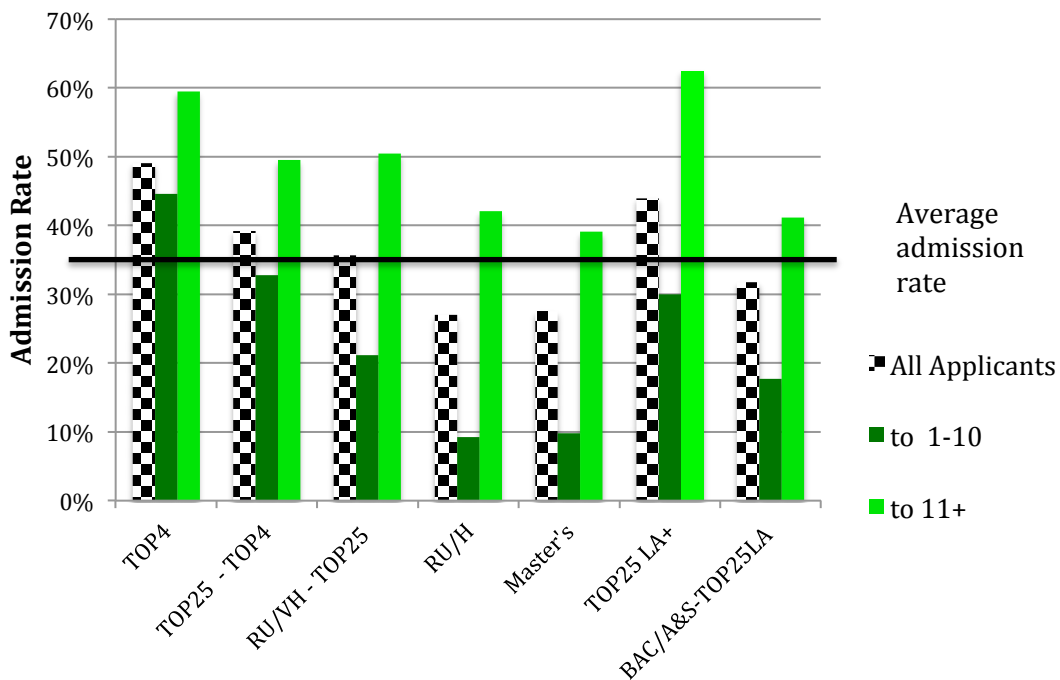


Figure 3.1: Percent of applicants from the seven “producer” groups accepted to all 14 graduate schools, those ranked 1-10, and those ranked 11+. The overall admission rate of 35% is indicated by the black horizontal line.

Figures 3.2 and 3.3 provide more detail for applicants to departments ranked 1-10. Figure 3.2 shows the number of students who applied, were admitted, and who ultimately matriculated to the consumer schools ranked 1-10.

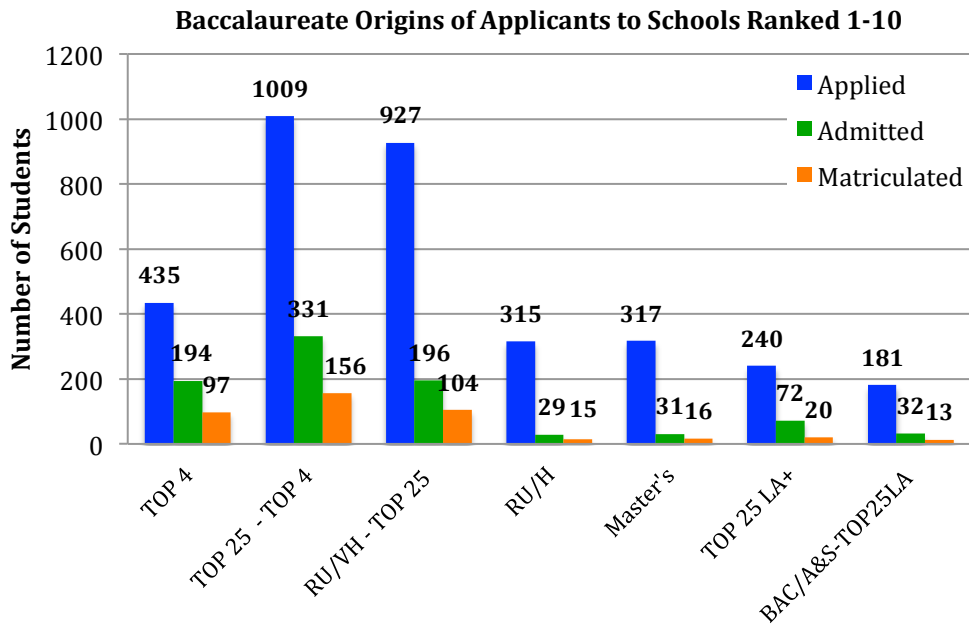


Figure 3.2: Total number of students applying, admitted, and matriculating to consumer schools ranked 1-10.

Figure 3.3 shows the same data by admission and matriculation rates. The overall admission rate for schools ranked 1-10 is 25% and the overall matriculation rate is 47%, as shown in Table 3.1.

Consider the TOP 25 LA+ category: Liberal Arts Colleges rated in the top 25 plus Olin College and Rose-Hulman. In total, 421 (out of 6640) applications came from TOP 25 LA+ institutions. For the departments ranked 1-10, 240 applications came from TOP 25 LA+ institutions. Of those 240 applications, 72 were admitted and 20 ultimately matriculated in the 1-10 departments. Hence, applicants from the TOP 25 LA+ producer group had an admission rate of 30% (slightly higher than the average) and a matriculation rate of 28% (significantly lower than the average).

The admission rate for students from the seven producer groups to departments ranked 1-10 ranges from 9% (for RU/H producer schools) to 45% (for producer schools in the TOP 4) is shown in Figure 3.3. The matriculation rate ranges from 28% (TOP 25 LA+) to 53% (for RU/VH - TOP 25).

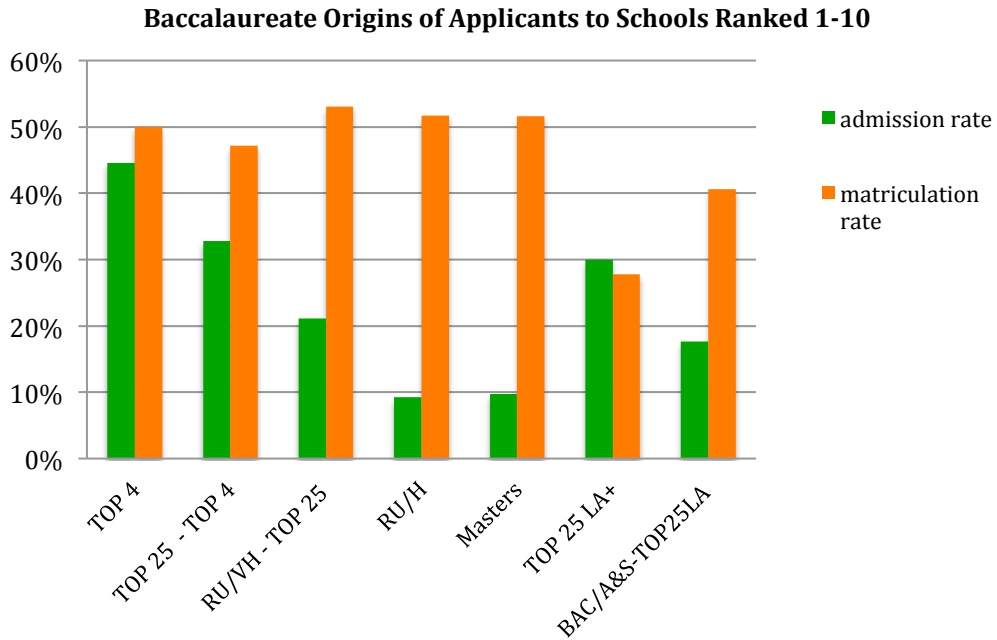


Figure 3.3: Admission and matriculation rates for consumer schools ranked 1-10.

Figures 3.4 and 3.5 are analogous to Figures 3.2 and 3.3 for schools ranked 11+. For the 11+ group, admission rates for the seven producer groups range from 39% (for master’s institution producer schools) to 62% (for the TOP 25LA+ group). Matriculation rates range from 21% (TOP 25 LA+) to 43% (for RU/H). It is notable that matriculation rates to 11+ consumer departments show different trends than for the schools ranked 1-10 and are particularly low for students from TOP 4 and TOP 25 LA+ institutions.

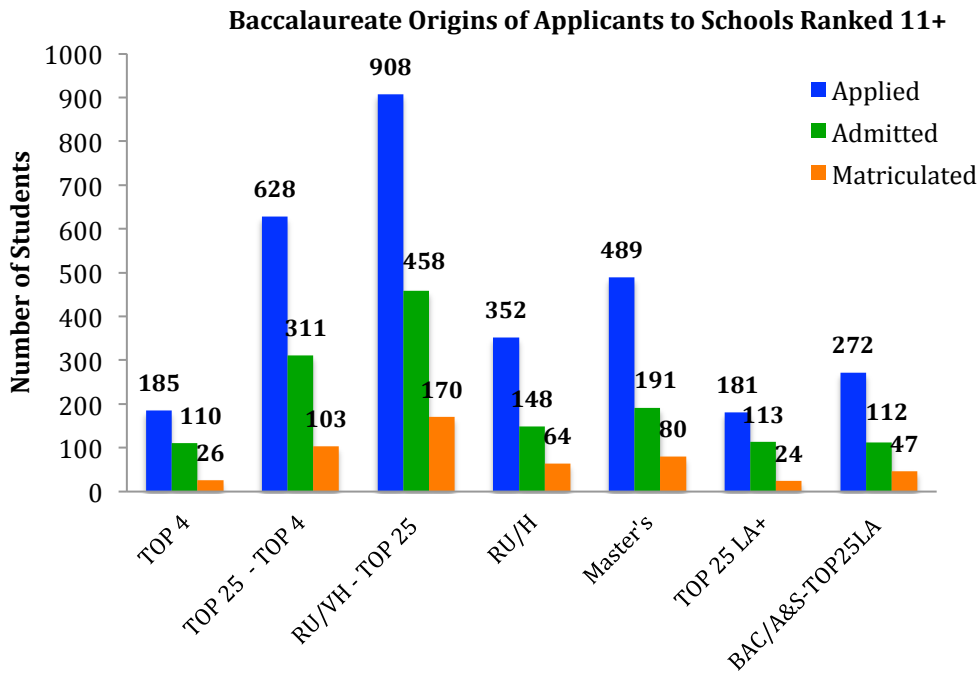


Figure 3.4: Total number of students applying, admitted, and matriculating to consumer ranked 11+.

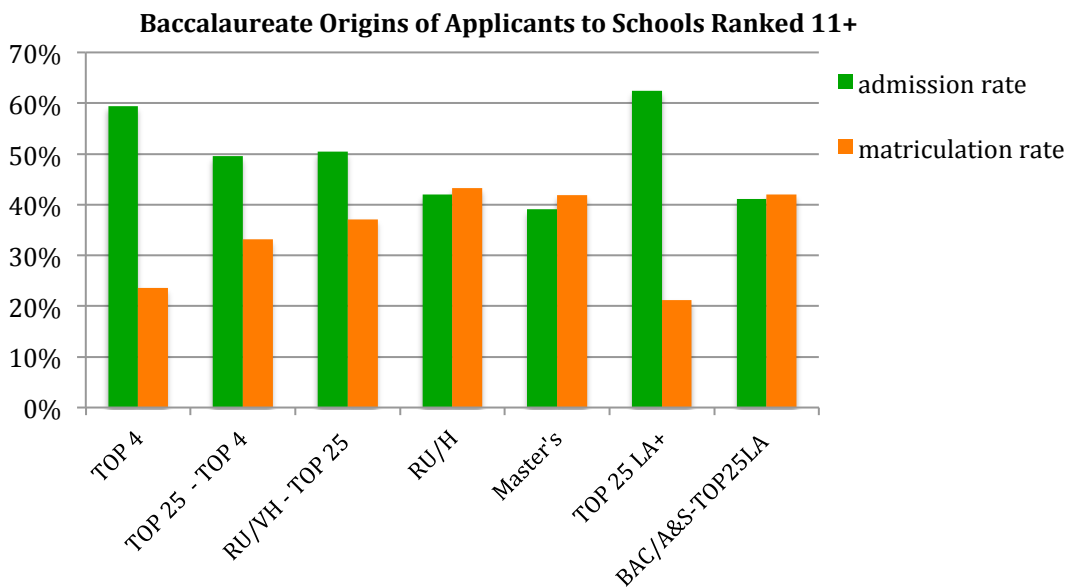


Figure 3.5: Admission and matriculation rates for consumer schools ranked 11+.

We conclude this section with some data refining the 14 departments a little further. Table 3.3 gives admission and matriculation rates when partitioning the 14 consumer departments into four consumer bins based on their rank.

Department rank	Average number of applications	Admission rate	Matriculation rate
1-10	918	25%	47%
11-20	537	42%	33%
21-30	230	38%	41%
31+	175	72%	40%

Table 3.3: Admission and matriculation rates for the 14 departments broken into 4 bins.

Each of the four bins contains 3 or 4 departments. Admission and matriculation rates for departments ranked 1-10 are as reported in Table 3.1. The ten departments ranked 11+ are grouped into three bins. Table 3.3 shows admission and matriculation rates for domestic applicants for each bin.

The average number of applications (column 2) is the total number of application records for the departments in a bin divided by the number of departments in the bin. Note that application records span the time period from 2007 to 2013, with not all departments providing records for all early and later years. Hence, it is difficult to make an accurate conclusion about the number of applications per year. To provide a perspective, Figure 3.6 shows for each of the 14 departments the number of domestic applications received and the number of domestic Ph.D. students enrolled in fall 2013 (applications were received for fall 2013 admission). The four darker circles represent the four departments ranked 1-10.

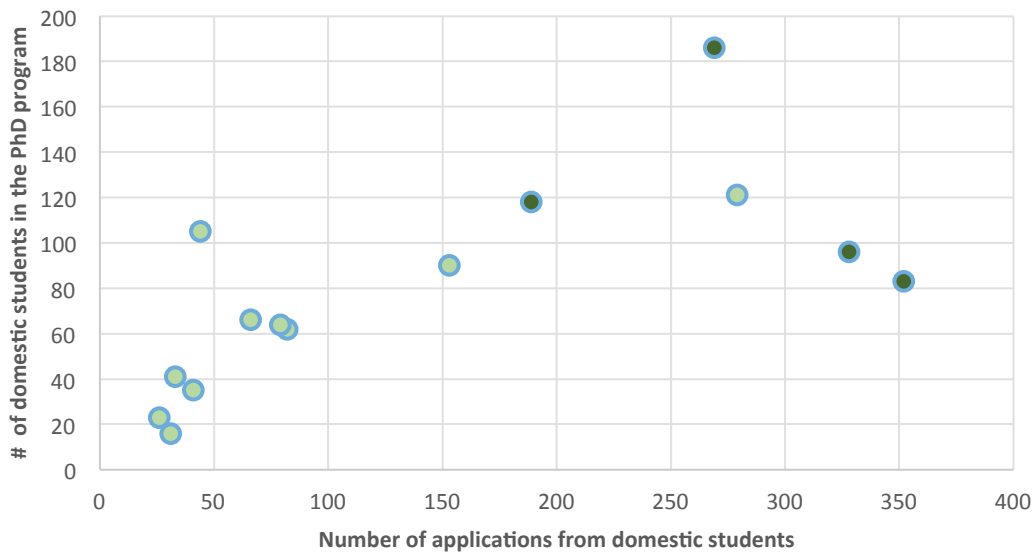


Figure 3.6: Number of domestic applications versus numbers of domestic students enrolled (Fall 2013).

For the 14 departments, the fall 2013 data shows that the percentage of applications from domestic students ranged from 6% to 31% and the percentage of domestic Ph.D. students enrolled students ranged from 21% to 46%.

4. Traditionally Underrepresented Groups

In this section we describe observed trends for applications identified as coming from females and two traditionally underrepresented groups, African-American and Hispanic/Latino students. Among all 7,032 application records, 99.4% report gender and 14% of all applicants reporting gender are female (983 applications).

A few departments did not provide ethnicity data on their applicants and 76% of the application records provide ethnicity. Departments reporting ethnicity did so in various ways, including ethnicity written in by the applicant or the applicant selecting from a predefined set of choices. For all application records reporting ethnicity, 2.93% are identified as African-American (155 applications) and 2.94% are identified as Hispanic/Latino (156 applications).

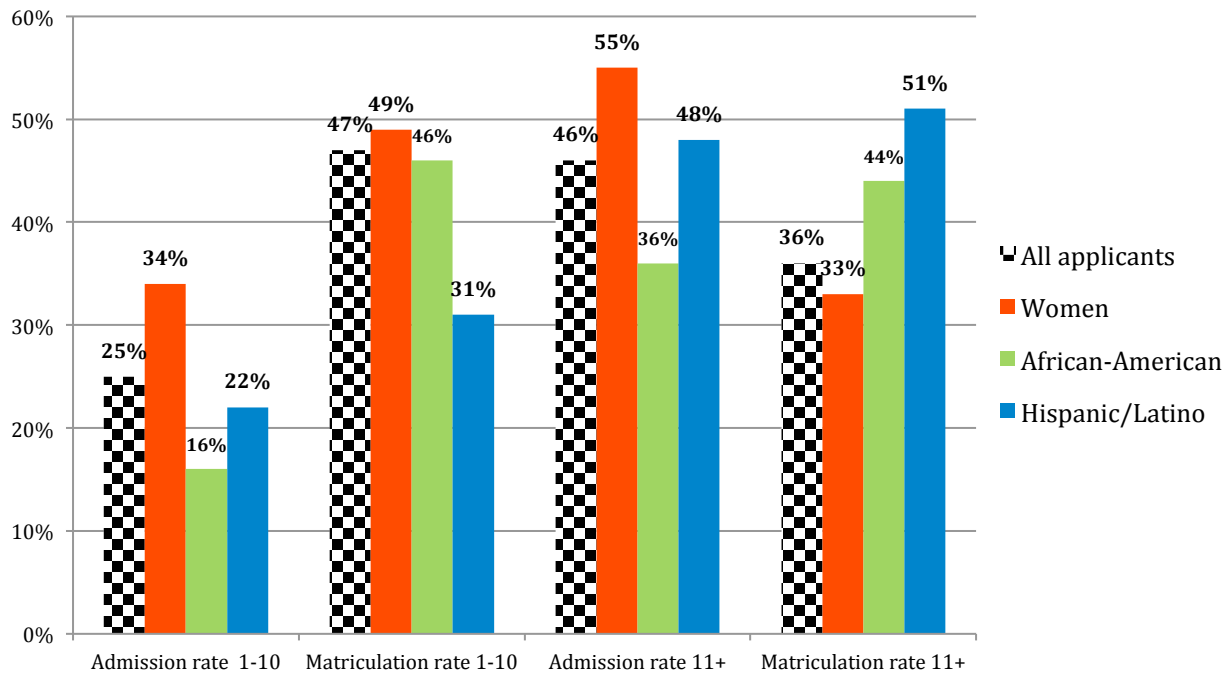


Figure 4.1: Percentage of students from traditionally underrepresented groups accepted and matriculating to schools ranked 1-10 and 11+.

Figure 4.1 shows admission and matriculation rates for women, African-Americans, and Hispanic/Latino students applying to consumer schools ranked 1-10 and ranked 11+. Recall that the admission rate represents the percentage of students admitted for each group and the matriculation rate is the percentage of the admitted students accepting admission and matriculating. The checkered bar represents the overall rates for

comparison. Other groups (including American Indian/Alaskan and Multiracial) make up less than 1% (52 applications) of the application records providing ethnicity.

Table 2.1 shows that 52% of all applications were to departments ranked 1-10. Looking at this breakdown for the three underrepresented groups, we see: 55% of the female applications, 52% of the African-American applications, and 38% of the Hispanic/Latino applications were made to departments ranked 1-10.

The data in Figure 4.1 show the following trends:

- Female applicants have higher overall averages (compared to the averages for all applicants) for both admission and matriculation rates in to departments ranked 1-10. For departments ranked 11+, the admission rate for woman is above average and the matriculation rate is slightly below average.
- For African-American applicants, the admission rate is significantly below the overall rates for both consumer groups. Matriculation rates are close to the average for departments ranked 1-10 and above average for departments ranked 11+.
- For Hispanic/Latino applicants, the admission rate is close to the overall rates for all students (below the average for 1-10 and above the average for 11+). However, the matriculation rate is significantly below average for Hispanic/Latino students accepted to departments ranked 1-10 and significantly above average for departments ranked 11+.

We point out that drawing further conclusions or inferences on the applications coming from African-American and Hispanic/Latino students is difficult due to the small number of applications. Information we consider interesting is a breakdown of all applications coming from underrepresented groups with respect to the seven producer groups as shown in Figure 4.2.

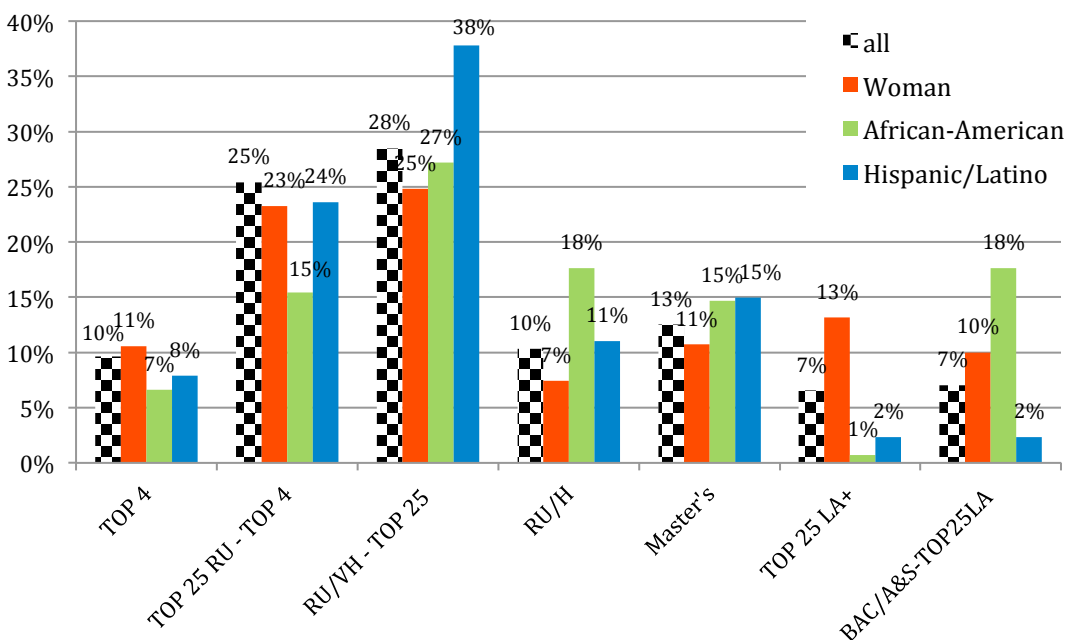


Figure 4.2: Breakdown of applications for each of the seven major producer groups by underrepresented groups.

The percentages shown for each of the four groups (all, Woman, African-American, Hispanic/Latino) shown in Figure 4.2 sum up to 100%. For example, 38% of all Hispanic/Latino applications came from students finishing their baccalaureate degrees at an RU/VH institution not ranked in the TOP 25.

The percentage of students in each group coming from the TOP 25 LA+ producer group is different from those of the other groups: The TOP 25 LA+ group has a higher percentage of female applicants and significantly lower percentages for African-American and Hispanic/Latino students applying to graduate school. Further study is needed to understand this phenomenon.

Only 22% of all African-American applicants come from an undergraduate institution with a CS department ranked in the TOP 25 group compared to 32% of all Hispanic/Latino applicants and the overall percentage being 35%.

5. GPA and GRE

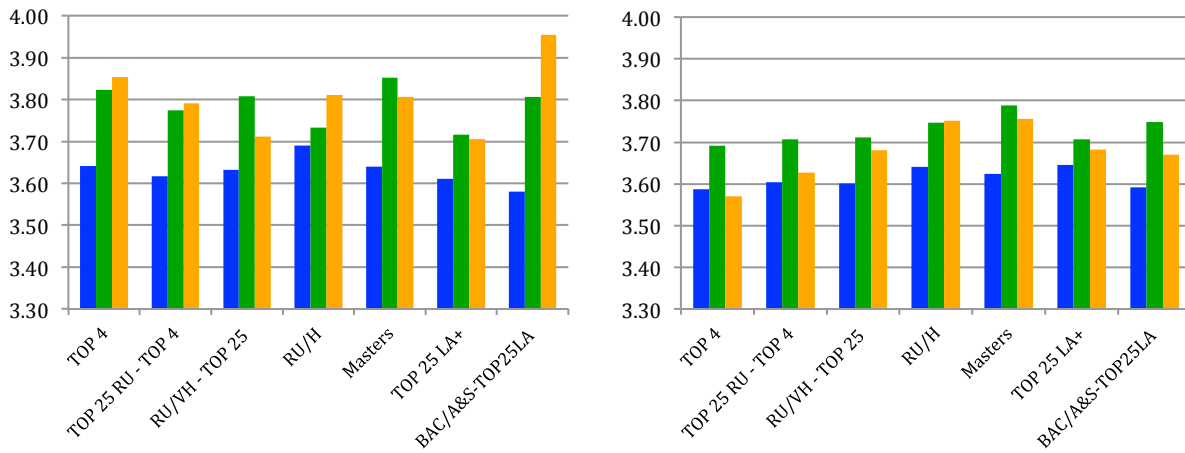
The majority of the departments provided GPA and GRE data of the applicants. Over 81% of the application records reported a GPA. We use only the 71% of records where GPA is reported on a 4-point scale.

Table 5.1 provides a summary for all departments, departments ranked 1-10 and departments ranked 11+. Overall, we find few surprises or major trends in the GPA data. The mean GPA of students applying to top 10 departments varies little from that of students applying to lower ranked departments. The difference in the mean GPA increases for admitted students and then again for matriculated students. The mean GPA for students matriculating in departments 11+ decreases compared to that of admitted ones, reflecting the challenge many departments face in attracting their best domestic applicants.

	All applicants		Admitted applicants		Matriculated applicants	
	GPA	Variance	GPA	Variance	GPA	Variance
All departments	3.62	0.32	3.75	0.25	3.74	0.27
1-10 departments	3.63	0.32	3.79	0.23	3.81	0.22
11+ departments	3.60	0.32	3.71	0.26	3.67	0.29

Table 5.1: GPA mean and variance for applicants and the consumer groups ranked 1-10 and 11+.

Figure 5.2 shows the mean GPAs for students from the seven producer groups and partitions the applications into the consumer groups 1-10 and 11+.



Applications to 1-10 departments

Applications to 11+ departments

Figure 5.2: Mean GPA for students in the seven producer groups applying (blue), admitted (green), and matriculating (orange) to 1-10 and 11+ departments.

Table 5.3 shows GPA mean and variance for three historically underrepresented groups in computer science: Females, African-Americans, and Hispanics/Latinos. Table 5.3 shows that while there is some range in the GPAs of applicants across the considered groups, the range among admitted students is very small. We believe that the convergence of the GPA profile of admitted students is relevant when advising students considering graduate school.

	All applicants		Admitted applicants	
	GPA	Variance	GPA	Variance
All applicants	3.62	0.32	3.75	0.25
Woman	3.61	0.32	3.73	0.27
African-American	3.41	0.36	3.69	0.31
Hispanic/Latino	3.54	0.36	3.66	0.29

Table 5.3: GPA mean and variance for all applicants and applicants from traditionally underrepresented groups.

	All applicants		Admitted applicants		Matriculated applicants	
	GPA	Variance	GPA	Variance	GPA	Variance
All females	3.61	0.32	3.73	0.27	3.71	0.30
Female 1-10	3.62	0.31	3.77	0.27	3.77	0.27
Female 11+	3.60	0.33	3.69	0.27	3.66	0.32

Table 5.4: Mean GPA and variance for female students applying to departments ranked 1-10 and 11+.

Table 5.4 shows the GPA information broken into female students applying to departments ranked 1-10 and 11+, respectively, and it also shows the GPA's of students matriculating. We do not show this breakdown for African-American and Hispanic/Latino applicants due to the small numbers in our pool. The trends are similar to what is shown in Table 5.1.

Finally, we give a brief summary of the GRE scores of the 72.4% of application records reporting GRE scores on the 800 scale (as opposed to the new 170 point scale).

	All applicants		Admitted applicants	
	GRE-Q	Variance	GRE-Q	Variance
All applicants	761	52.95	774	38.89
Woman	755	63.04	770	39.42
African-American	697	92.32	743	65.56
Hispanic/Latino	738	63.52	766	33.70

Table 5.5: GRE-Q mean scores and variance for all applicants and applicants with underrepresentation in the field on the 800 scale.

Tables 5.5 and 5.6 show quantitative and verbal GRE scores for all applicants as well as for females, African-Americans, and Hispanics/Latinos. These data indicate that variance in the applicant pool is considerably larger than the variance in the group of admitted students. This parallels the findings for GPAs.

	All applicants		Admitted applicants	
	GRE-V	Variance	GRE-V	Variance
All applicants	594	94.76	615	88.53
Woman	593	99.07	618	92.18
African-American	505	104.78	554	85.10
Hispanic/Latino	553	101.81	591	85.55

Table 5.6: GRE-V mean scores and variance for all applicants and applicants with underrepresentation in the field.

6. Recommendations and Conclusions

While the data collected for our admissions study shows some patterns, we caution that even with 14 participating schools and over 7,000 application records, we see only a limited view of the entire landscape of domestic applications to graduate schools. Nonetheless, we are confident in making several observations based on these data.

We offer a number of observations and recommendations to both “consumers” (graduate programs recruiting students) and “producers” (the institutions sending students to graduate school). Of course, many institutions are consumers and producers.

Observations for consumers:

- Students from master's institutions and RU/H schools are a significant source of applicants but are underrepresented in the pool of admitted students. Graduate committees should endeavor to understand programs at such institutions and build strategic partnerships for mutual benefit with some of them. Such partnerships will help departments identify promising students with potential of success.
- For graduate schools ranked 1-10, the admission rate from Top 25 LA+ schools is very high (second highest after Top 4), but the yield (i.e., matriculation rate) is lowest by a large margin. For 11+ schools, the admission rate from Top 25 LA+ schools is the highest among the 7 producer groups, but the yield is the lowest. The reasons are unclear, but we make one conjecture: students from TOP 25 LA+ schools are typically accustomed to a high level of personal communication. Increasing communication with top applicants may enhance the yield from this pool.
- Tables 5.3 to 5.6 show that female applicants from TOP 25 LA+ institutions are well prepared (based on GPA and GRE scores). Applications from female students from Liberal Arts institutions not in the top 25 also contain strong applications. While applications from the Liberal Arts institutions represent 14% of all application records, they represent 21% of all application records from females. A strategy for increasing the number of domestic female students in a department should take a careful look at the applicants from liberal arts institutions.

Observations for producers:

- Students from RU/H and master's institutions comprise a significant fraction of the applicant pool but have much lower admission rates than their peers at other types of institutions. Producers should think about how to prepare and advise those students for graduate school.

Recommendations for consumers:

- Maintain good historic data on your applicants, compare those data with the results reported here, and understand application, admission, and matriculation for producer groups relevant to your program. Use your departmental data to identify areas where admissions and recruiting practices might be refined.
- If an institution is not recruiting the quantity, quality, and diversity of domestic applicants it seeks, we recommend building partnerships with baccalaureate institutions that can serve as good "pipelines" of applicants to the program. The data in this study suggest that some graduate schools have built healthy relationships with strong 4-year colleges and master's institutions in their region and receive a significant number of applicants from those schools.
- Anecdotally, we have seen that some of the institutions in our study with relatively high matriculation rates also have effective recruitment strategies including well-organized and welcoming events for admitted students and proactive outreach from faculty to admitted students.

Recommendations for producers:

- Provide good advising for undergraduates interested in continuing on to graduate school. This advising can include information sessions on how to apply to graduate school and other resources for students (e.g., see the resources at CRA-E's website [5] and the Conquer website [6]). In addition, the results of this study show GPA information that can also be useful in advising.
- Keep good records on your students who apply to graduate school, where they are accepted, and where they matriculate. Use these data to help provide advice and guidance to your students.
- Forge relationships with some graduate programs that are likely to be good matches for your students. These can be geographically nearby graduate schools or even programs further afield. Invite a faculty member from a prospective partner school to visit your school to meet your faculty, learn about your program, meet your students, and give a recruiting talk. Over time, these relationships can lead to pipelines that will be mutually beneficial.

7. References

1. WebCASPAR, Data on science and engineering at U.S. academic institutions, <https://ncesdata.nsf.gov/webcaspar/>
2. Susanne Hambrusch, Ran Libeskind-Hadas et. al., [Exploring the Baccalaureate Origin of Domestic Ph.D. Students in Computing Fields](#), Computing Research News, January 2013.
3. US News and World Report, [Computer Science Rankings](#), 2014.
4. The Carnegie Classification of Institutions of Higher Education, <http://classifications.carnegiefoundation.org/>
5. Education Committee of the Computing Research Association (CRA-E), <http://cra.org/crae/>
6. CRA'E's Computer Science Undergraduate Research Conquer website, <http://cra.org/conquer>