# Leaving Careers in IT: Differences in Retention by Gender and Minority Status

Paula E. Stephan and
Sharon G. Levin

Andrew Young School of Policy Studies, Georgia State University
Department of Economics, University of Missouri-St. Louis

October 2004

This paper is drawn from a larger project supported by NSF (Grant #ELA 008995) that examined how recruitment and retention varies by gender and minority status in the IT workforce. The full report is available at http://www.gsu.edu/~ecopes/itworkforce/reports/it_final_report.pdf.
The authors would like to thank Lawrence Burton, of Science Resources Statistics, National Science Foundation, for his assistance with the SESTAT data, especially with regard to the definition of IT fields and occupations. We also wish to thank Ephraim R. McLean, Department of Computer Information Systems, Georgia State University, for his input and advice. Grant Black provided graduate research assistance on this project when he was a PhD student at Georgia State University. At the University of Missouri-St. Louis, assistance was provided by Maria Tiratsuyan and Emily Trevathan, who was supported by a REU supplemental grant. Support for this project was also provided by the Andrew Young School of Policy Studies, Georgia State University and the Office of Research Administration, the University of Missouri-St. Louis.

There has been considerable interest in recent years concerning the low prevalence of women and underrepresented minorities in the information technology (IT) workforce. Traditionally this focus was motivated by concerns regarding equity. This interest was augmented during the 1990s because of the key role the IT sector played in the economic expansion and the concern that a shortage of IT workers existed. Increased participation of women and underrepresented minorities was seen as one way by which the IT workforce could be "grown." The focus of much of this discussion concerned how the pipeline leading to careers in IT could be expanded, making careers in IT more attractive and accessible for women and minorities. A case in point is the Carnegie Mellon initiative, "unlocking the clubhouse door," which focused on recruiting and attracting women and minorities, with considerable success, into IT programs at Carnegie Mellon University (Margolis and Fisher 2002).

The size of the IT workforce, of course, depends not only on the pipeline in; it also depends on the pipeline out. If individuals leave the IT workforce for another occupation, or leave to exit the labor force, the size of the IT workforce is diminished. The major focus of this study is to examine factors related to retention and how retention varies by gender and underrepresented minority status.

The focus of this study differs from that of most studies concerning women and underrepresented minorities in the IT workforce or, more generally, in the STEM (science, technology, engineering and math) workforce. Most articles look at why women and underrepresented minorities leave STEM fields while students; few examine retention after the career has begun, as we do here. Preston (2004) and Bentsen (2000) are two exceptions.

Preston, who focuses on women in science and engineering, finds that mentors play an important role in whether women exit their field of training after starting their career. Bensten, who examines IT careers specifically, focuses on policies directed at retention. She is particularly struck by the National Security Agency's (NSA) ability to retain the women that they hire. She quotes Bernard Norvell, an NSA technical director for Human Resources Services, as saying: "We can appeal to women with continuing-ed programs, onsite child-and elder-care responsibilities and fitness centers. These things are big sellers, very enticing to women." (Bensten 2000, p. 8). [1]

The importance of retention in determining the size and mix of the IT workforce can be illustrated by examining the composition of the 1999 IT workforce as measured by SESTAT and determining the shrinkage that occurred between 1993 and 1999. We estimate that the female IT workforce in 1999 would have been 28,000 larger if women trained in IT and working in IT in 1993 had been retained in IT in 1999; 83,000 larger if women who had worked in IT without formal training in IT in 1993 were working in IT in 1999. Combined this would have meant that the female IT workforce would have been 40 percent larger in 1999 than it was.[2] By comparison, the male IT workforce would have been about a fourth larger if men working in IT occupations in 1993 had been retained in IT occupations in 1999. Retention would have increased the underrepresented minority IT workforce by about 50 percent in 1999. Taken

[1] The Report of the Congressional Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development makes a similar point (p. 52) when it states: "Issues of work and family balance are of paramount concern to women employed in SET fields. Policies and benefits such as extended parental leaves, part-time work options, on-site child care facilities, and greater scheduling flexibility are needed. Access to quality, reliable, affordable child care is key to the retention of working parents."

[2] We estimate that the 1999 female IT workforce, as measured by SESTAT, was about 278,000, which includes 206,669 who were retained and 72,000 who were recruited into the IT workforce between 1993 and 1999. The 40 percent calculation uses 1999 as the base; the percent increase would have been about 35 percent if the 1993 base had been used instead.

together, these differential retention rates mean that the IT workforce would have been more representative of the U.S. population in 1999 than it was.

Statistics for the study come from the Scientists and Engineers Statistical Data System (SESTAT) compiled by the National Science Foundation.[3] The SESTAT data integrates three separate surveys overseen by SRS of individuals who have a college degree or higher who are either working in a science and engineering occupation or trained in science and engineering. The three surveys are: the Survey of Doctorate Recipients, the National Survey of College Graduates and the National Survey of Recent College Graduates.

This paper is organized into 5 parts. In section II we briefly describe the SESTAT database as well as our definition of what it means to be IT trained or working in an IT occupation. Section III describes the IT workforce, first, in terms of formal training in IT, including estimates of the size of the IT workforce in 1993 and 1999 and second, in terms of who is employed in IT occupations. We are particularly interested in drawing inferences concerning retention of the IT workforce and how these patterns differ by gender and minority status. The econometric analysis is presented in Section IV where the issue of retention of the IT trained is studied using a multinomial approach. Particular attention is paid to how retention varies by gender and underrepresented minority status. Conclusions are drawn in Section V.

## II. The SESTAT Database

The SESTAT data integrates three separate surveys overseen by the Science Resources Statistics (SRS) unit of the National Science Foundation: the Survey of Doctorate Recipients (SDR), the National Survey of College Graduates (NSCG) and the National Survey of Recent

[3] SESTAT is discussed in Section II of this report and details can be found at http://sestat.nsf.gov.

College Graduates (NSRCG). Although it is the best available database for the research we propose, SESTAT has several shortcomings. First, as is true with other databases, the SESTAT definition of IT occupations fails to capture all jobs where IT work is occurring. Second, SESTAT underrepresents four groups of scientists and engineers in the United States in 1995 and subsequent years. These are: (a) new immigrants with science and engineering (S&E) degrees earned outside the U.S. who entered the U.S. after 1990 and did not subsequently receive a degree in the U.S.; (b) college grads without S&E degrees who were not working in S&E occupations in 1993 but were in S&E occupations at a later date; (c) associate degree holders working in the S&E workforce; (d) individuals who lack any formal degree but who are working in the S&E workforce. In addition, no one over the age of 75 is included in the sample. A third shortcoming of SESTAT is that individuals without S&E training who began working in IT occupations in 1993 or later are not included. Fourth, and of importance for this study, programming, both as an occupation, and as a field of education, is not defined by SESTAT as being in S&E. This does not mean that programmers are excluded from SESTAT. It does, however, mean that they are not intentionally counted by SESTAT. Thus, individuals who worked as computer programmers in 1993 are only included in SESTAT if they received a degree in a S&E field and individuals who were trained in programming are only included in SESTAT if they were working in an S&E occupation in 1993. A final limitation of the data is that degrees awarded from business schools are excluded from the definition of S&E degrees regardless of content. For our purposes, this means that degrees in computer and information

sciences awarded by business schools are missed in our definition of formal training.[4] Table 1 summarizes the content of the SESTAT database.[5]

**Table 1**. The SESTAT Database

| Survey | 1993 | 1995 | 1997 | 1999 |
|---|---|---|---|---|
| National Survey of College Graduates (NSCG) | All individuals identified as having a S&E degree in 1990 Census; All individuals identified as having a non-S&E college degree in 1990 who hold an S&E occupation in 1993. All individuals must be living in the U.S. U.S. earned doctorates are excluded. | All individuals in the 1993 NSCG; Individuals are added if they received an S&E degree between 1990 and 1994. U.S. earned doctorates are again excluded. | All individuals in the 1993 NSCG; Individuals are added if they received an S&E degree between 1990 and 1996. U.S. earned doctorates are again excluded. | All individuals in the 1993 NSCG; Individuals are added if they received an S&E degree between 1990 and 1998. U.S. earned doctorates are again excluded. |
| National Survey of Recent College Graduates (NSRCB) | Individuals who earned bachelor's or masters S&E degrees in May to December of 1990 or academic years 1991 or 1992. | Individuals who earned bachelor's or master's S&E degrees in academic years 1993 or 1994 | Individuals who earned bachelor's or master's S&E degrees in academic years 1995 or 1996. | Individuals who earned bachelor's or master's S&E degrees in academic years 1996 or 1997. |
| Survey of Doctorate Recipients (SDR) | Individuals who earned S&E doctorates in U.S. through academic year 1992 and indicated they planned to stay in the U.S. at time degree was received. | Individuals who earned S&E doctorates in U.S. through academic year 1994 and indicated they planned to stay in the U.S. at time degree was received. | Individuals who earned S&E doctorates U.S. through academic year 1996 and indicated they planned to stay in the U.S. at time degree was received. | Individuals who earned S&E doctorates in U.S. through academic year 1998 and indicated they planned to stay in the U.S. at time degree was received. |

[4] This may explain why we find that about 10% of those working in IT occupations without a formal IT degree have a degree in business administration and management.

[5] SESTAT is discussed in several places. http//srsstats.sbe.nsf.gov/techinfo.html has a document called "Design and Methodology." Also, NSF has published SESTAT: A Tool for Studying Scientists and Engineers in the U.S.

## Defining IT Trained and IT Occupations using SESTAT

Defining IT training and IT occupations is not a straightforward task. This is because IT skills are often learned as part of a larger educational program or because for many older workers IT degrees were not awarded when they were in school. It is also because many individuals learn IT skills on the job, eschewing any formal educational training. Recently, two reports have drawn on data from SESTAT to analyze the IT workforce. The first of these is Building a Workforce for the Information Economy (National Research Council); the second is the IT Data Project (Ellis and Lowell). We draw heavily on these reports in defining IT occupations as well as IT training.

We consider individuals to be formally trained in IT if they received one or more degrees in

- Computer/information sciences
- Computer science
- Computer system analysts
- Information service and systems
- Other computer and information sciences
- Computer and system engineers
- Electrical, electronics and communications engineering if the recipient also had minored or did a second major in areas of study in computer or information sciences.

We consider individuals to be working in an IT occupation if they are

- A computer analyst
- Computer scientists, except system analysts

- Information system scientists and analysts
- Other computer and information science occupations
- Computer engineers; software engineers
- Post-secondary teachers in computer and mathematical sciences
- Computer engineers—hardware
- Computer programmers[6]

## III. Retention Six Years Later

The primary focus of this research is whether individuals who are working in IT occupations in one time period are working in IT occupations at a later date. We focus on two distinct groups of IT workers: those with formal training in IT and those who are not formally trained in IT. Persistence is measured by whether the individual is retained in an IT occupation in the six year interval from 1993 to 1999.

Descriptive statistics are presented in Table 2 for the 1,058,989 1993 IT workers (weighted cases) who responded to both the 1993 and 1999 surveys. We find that the IT workforce was primarily white (83.6%) and male (70.3%). Before discussing the retention results, we note that informal training plays an important role in drawing people into the IT workforce. To wit, by inference we see that over 60% of the 1993 IT workforce was classified

[6] Recall that programming was not included as a S&E field of training nor occupation in SESTAT so that the only programmers included in SESTAT are those who were trained in an S&E field and work as programmers or individuals not trained in S&E but working in S&E occupations in 1993.

as not having formal IT training, using the above definitions.[7] We also note that the proportion without formal training is slightly higher among women than it is among men.

**Table 2.** Weighted means for individuals employed in IT occupations in 1993 and in SESTAT in 1999.

| | All | Females | Males | Whites | Asians | African Americans | Hispanic & Others |
|---|---|---|---|---|---|---|---|
| Ittrain93 | 0.387 | 0.366 | 0.395 | 0.384 | 0.571 | 0.454 | 0.387 |
| retained | 0.710 | 0.658 | 0.732 | 0.703 | 0.790 | 0.660 | 0.716 |
| retained & IT trained | 0.804 | 0.735 | 0.824 | 0.800 | 0.840 | 0.778[a] | |
| retained & not IT trained | 0.651 | 0.604 | 0.672 | 0.649 | 0.725 | 0.618[a] | |
| wknoit99 | 0.232 | 0.247 | 0.225 | 0.236 | 0.155 | 0.316 | 0.237 |
| nowork99 | 0.059 | 0.095 | 0.043 | 0.061 | 0.055 | 0.025 | 0.047 |
| unempl99 | 0.012 | 0.011 | 0.013 | 0.012 | 0.006 | 0.018 | 0.015 |
| outlf99 | 0.046 | 0.084 | 0.031 | 0.049 | 0.049 | 0.007 | 0.032 |
| n | 1,058,989 | 314,564 | 744,425 | 885,600 | 97,688 | 44,914 | 30,786 |
| % of sample | 100.0% | 29.7% | 70.3% | 83.6% | 9.2% | 4.2% | 2.9% |

Notes: Others include Native Americans. ittrain93=1 if trained in IT discipline in 1993; retained=1, if employed in IT occupation in 1999; wknoit99=1, if employed in a non-IT occupation; nowork99=1, if not working in 1999; unempl99=1, if unemployed in 1999; outlf99=1, if out of the labor force in 1999. [a]These means are for all underrepresented minorities combined: African Americans and Hispanic and Others.

[7] The role of recruitment is especially notable given that the SESTAT database does not define programming, either as an occupation or as a field of training, as being in S&E. Undoubtedly, recruitment would play an even larger role if our analysis included those working as programmers who did not have formal training in a SESTAT defined S&E field or who had been working in another S&E occupation.

Retention among the million plus 1993 IT workers is approximately seven in ten. About a quarter of those who were in IT in 1993 are now in a non-IT occupation; approximately 6 percent are no longer working. Most of these are out of the labor force, but a small percent are unemployed. We see that whites had a higher retention rate (70.3%) than African Americans (66.6%) but the retention rate for Hispanics was slightly higher (71.6%) than that for whites, while that for Asians was considerably higher (79.0%). The differential in retention rates by gender is striking: 73.2% of men who were working in IT in 1993 are working in IT in 1999 while 65.7% of the women who were working in IT are in IT in 1999. The differential is explained in part by a higher proclivity of women to move from IT positions to not working; the movement of women to non-IT positions is somewhat comparable to that of men. By way of contrast, the lower retention rate for African Americans compared to whites reflects a movement into non-IT occupations, not a movement into the "no work" category.

Retention also varies by training status: those with formal training in IT are considerably more likely to remain in IT than those without formal training (80.4% vs. 65.1%). This pattern holds across gender and minority status.

**Section IV. Multinomial Analysis of Retention**

Here we examine the factors explaining the odds that those working in IT in 1993 are retained in IT occupations in 1999.

Three possible outcomes can be observed for each individual in 1999: (1) to remain in an IT occupation (WORK IT); (2) to work in a non-IT occupation (WORK NOT-IT); and (3) to not work (NO WORK). Given these three possibilities, the estimation procedure we use is multinomial logit. Econometric theory tells us that from an efficiency standpoint it is preferable

to estimate a multinomial logit model than a series of separate binomial logit models comparing the log odds of one outcome versus another (Allison, 1999, p. 123).

For ease of exposition, we present not only the actual logit coefficients from the estimated models[8] but also the more easily interpreted odds-ratio. By way of example, for a dummy variable such as female, a value of 1.0 of the odds-ratio indicates that the odds of the event in question occurring are the same for women as for men, the benchmark. An odds-ratio greater than 1.0, for example, 1.5, tells us that the odds of the event in question occurring for women are 1.5 times those for men. (In other words, the odds are 50% higher for women than for men). An odds-ratio less than 1.0, for example, 0.4, indicates that the odds for women are just 0.4 times the odds for men (or that the odds for men are 1.0/0.4 or 2.5 times the odds for women.)[9] The odds-ratios are computed by taking the exponent of the log-odds for each variable. We also indicate whether each parameter estimate is statistically significant and whether we can reject the null hypothesis that all the parameters in the model have no affect on the odds-ratios. This is indicated by the statistical significance of the likelihood-ratio test ( -2 Log L). A glossary of the variables used in the models estimated appears in the Appendix.

Table 3 presents the sample statistics. The logit results are presented in Table 4. In order to demonstrate the increased insight gained by using the mutinomial approach, we first present a simple logit equation, estimating the odds that an individual remains in an IT occupation vs. leaves an IT occupation, either to work in another occupation or to not work, being either out of the labor force or unemployed. These logit results are presented in column 1 of Table 4.

[8] The logit coefficient shows by how much the log-odds of one outcome compared to another varies as each explanatory variable changes by one unit, holding all else constant.
[9] For a quantitative variable, if we subtract 1 from the odds-ratio and multiply by 100, the resulting number can be interpreted as the percent change in the odds for each unit increase in the independent variable.

**Table 3.** Sample statistics

| | All | | IT trained | | Non-IT trained | |
|---|---|---|---|---|---|---|
| | n=5208 | | n=2110 | | n=3098 | |
| Variables | Mean | Std. | Mean | Std. | Mean | Std. |
| ittrain93 | 0.405 | 0.491 | | | | |
| itlatest93 | 0.383 | 0.486 | 0.945 | 0.227 | | |
| female | 0.269 | 0.443 | 0.279 | 0.449 | 0.261 | 0.440 |
| african american | 0.043 | 0.202 | 0.046 | 0.210 | 0.040 | 0.196 |
| asian | 0.149 | 0.356 | 0.199 | 0.399 | 0.115 | 0.319 |
| hispanothers | 0.043 | 0.203 | 0.037 | 0.190 | 0.047 | 0.212 |
| perm93 | 0.112 | 0.316 | 0.131 | 0.338 | 0.099 | 0.299 |
| temp93 | 0.092 | 0.289 | 0.130 | 0.337 | 0.066 | 0.249 |
| age93 | 38.559 | 8.068 | 35.639 | 6.816 | 40.548 | 8.248 |
| agesq93 | 1551.9 | 647.312 | 1316.57 | 515.743 | 1712.18 | 678.178 |
| addit | 0.009 | 0.094 | 0.009 | 0.092 | 0.009 | 0.095 |
| addnoit | 0.032 | 0.176 | 0.034 | 0.182 | 0.030 | 0.172 |
| gotmarried | 0.106 | 0.308 | 0.124 | 0.329 | 0.094 | 0.292 |
| gotsingle | 0.041 | 0.199 | 0.045 | 0.208 | 0.038 | 0.191 |
| gotchildren06 | 0.146 | 0.353 | 0.182 | 0.386 | 0.122 | 0.327 |
| gotperm | 0.042 | 0.201 | 0.058 | 0.233 | 0.031 | 0.174 |
| selfemp | 0.05 | 0.219 | 0.036 | 0.185 | 0.061 | 0.239 |
| othersci | 0.192 | 0.394 | | | 0.323 | 0.468 |
| othereng | 0.142 | 0.349 | | | 0.239 | 0.426 |
| bus | 0.100 | 0.300 | | | 0.168 | 0.374 |
| socsci | 0.054 | 0.226 | | | 0.091 | 0.288 |
| fgotmarried | 0.027 | 0.163 | 0.029 | 0.169 | 0.026 | 0.160 |
| fgotsingle | 0.014 | 0.118 | 0.017 | 0.130 | 0.012 | 0.109 |
| fgotchildren06 | 0.041 | 0.199 | 0.056 | 0.230 | 0.032 | 0.175 |
| Work It | 0.735 | 0.441 | 0.814 | 0.389 | 0.682 | 0.466 |
| Work Not-It | 0.209 | 0.406 | 0.143 | 0.350 | 0.254 | 0.435 |
| Not Work | 0.056 | 0.230 | 0.044 | 0.204 | 0.064 | 0.245 |

Note: Variables indicated by "f" times the variable denote an interaction with the female dummy variable.

**Table 4**. The odds of different employment outcomes in 1999 for those employed in IT occupations in 1993

| | Using Logit | | Using Multinomial Logit | | | | | |
|---|---|---|---|---|---|---|---|---|
| | WORK IT vs. NOT WORK IT | | WORK IT vs. WORK NOT-IT | | WORK IT vs. NO WORK | | WORK NOT-IT vs. NO WORK | |
| | **(1)** | | **(2)** | | **(3)** | | **(4)** | |
| Variables | Coeff. | Odds ratio | Coeff. | Odds ratio | Coeff. | Odds ratio | Coeff. | Odds ratio |
| intercept | ***-3.526*** | 0.029 | -1.041 | 0.353 | ***-4.526*** | 0.011 | **-3.485** | 0.031 |
| Ittrain93 | 0.274 | 1.316 | 0.323 | 1.381 | -0.249 | 0.780 | -0.348 | 0.706 |
| itlatest93 | ***0.744*** | 2.104 | ***0.822*** | 2.275 | 0.480 | 1.616 | -0.342 | 0.710 |
| female | ***-0.243*** | 0.785 | -0.070 | 0.932 | ***-0.853*** | 0.426 | ***-0.783*** | 0.457 |
| african american | -0.253 | 0.776 | **-0.360** | 0.698 | 0.420 | 1.522 | **0.779** | 2.179 |
| asian | 0.178 | 1.195 | 0.231 | 1.260 | -0.071 | 0.931 | -0.302 | 0.739 |
| hispanothers | 0.050 | 1.051 | 0.054 | 1.055 | 0.112 | 1.119 | 0.058 | 1.060 |
| perm93 | 0.052 | 1.053 | 0.050 | 1.051 | 0.182 | 1.200 | 0.132 | 1.141 |
| temp93 | 0.296 | 1.344 | 0.268 | 1.307 | 0.523 | 1.687 | 0.255 | 1.290 |
| age93 | ***0.216*** | 1.241 | 0.070 | 1.073 | ***0.452*** | 1.571 | ***0.382*** | 1.465 |
| agesq93 | ***-0.003*** | 0.997 | -0.001 | 0.999 | ***-0.007*** | 0.993 | ***-0.006*** | 0.994 |
| addit | 0.668 | 1.950 | 0.527 | 1.694 | *na* | | *na* | |
| addnoit | ***-1.016*** | 0.362 | ***-1.119*** | 0.327 | -0.323 | 0.724 | 0.796 | 2.217 |
| gotmarried | 0.129 | 1.138 | 0.105 | 1.111 | 0.171 | 1.186 | 0.066 | 1.068 |
| gotsingle | -0.006 | 0.994 | 0.043 | 1.044 | -0.280 | 0.756 | -0.323 | 0.724 |
| gotchildren06 | -0.048 | 0.953 | -0.020 | 0.980 | 0.027 | 1.027 | 0.047 | 1.048 |
| gotperm | -0.389 | 0.678 | -0.411 | 0.663 | -0.322 | 0.725 | 0.090 | 1.094 |
| selfemp | ***-0.446*** | 0.640 | ***-0.397*** | 0.672 | ***-0.675*** | 0.509 | -0.278 | 0.757 |
| othersci | ***0.632*** | 1.881 | ***0.632*** | 1.881 | ***0.555*** | 1.742 | -0.078 | 0.925 |
| othereng | **0.260** | 1.297 | 0.211 | 1.235 | **0.577** | 1.781 | 0.366 | 1.442 |
| bus | 0.054 | 1.055 | 0.014 | 1.014 | 0.202 | 1.224 | 0.188 | 1.207 |
| socsci | 0.199 | 1.221 | 0.137 | 1.147 | 0.362 | 1.436 | 0.225 | 1.252 |
| fgotmarried | -0.237 | 0.789 | -0.262 | 0.770 | -0.291 | 0.748 | -0.029 | 0.971 |
| fgotsingle | 0.615 | 1.849 | 0.548 | 1.730 | 0.854 | 2.349 | 0.306 | 1.358 |
| fgotchildren06 | -0.087 | 0.916 | 0.285 | 1.330 | ***-1.162*** | 0.313 | ***-1.447*** | 0.235 |
| n | 5208 | | 5208 | | | | | |
| -2 log L | ***5721.1*** | | ***6854.7*** | | | | | |

Notes: Estimates indicated by ***XXX*** are statistically significant at the .01 level, **XXX** at the .05 level, and XXX at the .10 level. *na* means that the parameter could not be estimated because of the lack of variation within the category. Variables indicated by "f" times the variable name denotes an interaction with the female dummy variable.

We find strong evidence that retention relates to training, but the story is somewhat nuanced. There is no evidence that across the board those with one or more IT degrees are more likely to be retained than those without an IT degree. Rather, retention relates to whether the IT

degree is the latest degree that the individual had in 1993. Thus, while those with an undergraduate degree in computer science but an MBA are not significantly more likely to be retained, than those with an undergraduate degree in business but a masters in computer science are. And the odds are of a striking magnitude: those whose IT degree is their latest degree are more than two times as likely to be retained as those who are IT trained but have received a subsequent, non-IT degree. We also find that those who took additional training of a non-IT nature during the 1990s', arguably with the goal of changing careers, are .36 as likely to be retained as those who did not. Surprisingly enough, we find that taking additional IT training during the six year interval is not significantly related to retention. This may be due to the strong degree of multicolinearity between "itlatest93" and "addit." Among those without formal IT training who were working in IT in 1993, individuals trained in "other science" and "other engineering" (to distinguish them from training in IT) are significantly more likely to be retained than the benchmark group (education, humanities and the health professions).

Age (age93) increases the odds of being retained in IT but the odds increase more slowly as one ages, as shown by the negative age-squared term (agesq93). We find no evidence that retention is related to citizenship status, but we do find that retention is related to a change in citizenship status. Those who became permanent residents during the six year period are significantly less likely (at the ten percent level) to be retained than natives. This may reflect their freedom from H-1B visa status (and holding an IT job) upon gaining permanent residency in the United States.

For our purposes, we are particularly interested in the effects that gender and race/ethnicity have on retention, as well as family status variables that may differentially affect retention by gender and minority status. We find differences by gender to be statistically significant, other things being equal. Indeed, the odds that a woman remains working in IT are .785 those of a man remaining in IT. This is an important result in the sense that the introduction

of other variables does not "explain away" the gender effects we saw above in the descriptive data. In terms of minority status, we find that African Americans are .78 as likely to be retained in an IT occupation as are whites (the benchmark). The coefficient is significant at the 10 percent level. We find no evidence that retention, other things being equal, differs between the Hispanic and other group, and whites; neither do we find any evidence that retention significantly varies between Asians and whites, other things being equal.

We are able to observe several changes with regard to family status during the period. Specifically, we can determine if the respondent got married, became single or began to parent a child six or younger during the period. We enter all three change variables into the analysis. We find no statistical evidence that those who changed marital status or parenting status were any less likely to be retained than those who did not change status. In order to examine whether the effects differ by gender, we interact "female" with the change in status variables. We find that at the 10% level of significance, women IT workers who became single during the period are more likely to be retained than men who became single during the period. We find no evidence that for either men or women the addition of a young child in the household affects retention.[10]

Finally, we comment on the variable "selfemp" which measures whether the individual was working for her or him-self as an IT worker in 1993. Our priors are that retention is related to self employment, the notion being that self employment provides more flexibility. We find no statistical support for this prior. Indeed, instead we find that the self employed were less likely to be retained.

The right hand panel of Table 4 presents the multinomial logit results comparing three options, rather than one: the odds of staying in an IT position relative to leaving and taking a position in a non-IT occupation (Work IT vs. Work Not-IT); the odds of remaining in an IT

[10] The number of underrepresented minorities is too small to estimate the model interactively by minority status.

occupation vs. not working (Work IT vs. No Work); and the odds of leaving an IT occupation to work in a non-IT occupation vs. leaving an IT occupation to not work.

The results clearly indicate that increased insight can be gained when the choices are modeled as multiple outcomes vs. either or. This is especially the case for the gender and race/ethnicity variables. For example, we find that African Americans are significantly less likely to remain in IT vs. moving to a different occupation (column 2). Indeed, the odds that they stay are .70 those of whites (column 2). The odds of remaining in IT vs. leaving the work force are not significantly related to being African American, but the odds of working in a non-IT occupation vs. not working are significantly higher for African Americans. To wit, compared to whites, African Americans are more than twice as likely to be working in non-IT occupations as not working. Stated differently, we find evidence that blacks leave IT for another occupation; they do not leave IT to leave the labor force or because they are unemployed. Indeed, other things equal, they are much more likely than whites to be working in a non-IT occupation than to not work.

By way of contrast, in the multinomial logit model we find that gender is insignificantly related to remaining in IT vs. switching to another occupation (column 2); but women are significantly more likely to not work, regardless of whether the benchmark is remaining in IT vs. not working (column 3) or working in another occupation vs. not working (column 4). And the effects are reasonably large with women being only about .45 percent more likely to stay in the work force than men (or men being about 2.2 times more likely to stay than women). We also find evidence that adding young children to the family has differential effects by gender; women who began parenting young children during the period are significantly more likely to not be working, whether it is in IT or a non-IT occupation than are men who began parenting during the period. The effects are quite strong, with women being between .23 and .31 less likely to be working than men who added young children to their families. Clearly, the work force status of female IT workers is more sensitive to the arrival of children than is the workforce status of men.

The reader is reminded that the results presented above are conditional on the individual being in an IT occupation in 1993, regardless of whether the individual was IT trained or not IT trained. In order to ascertain whether these results vary by training status, we re-estimate the model for those with IT training (Table 5) and those without IT training (Table 6). The results suggest that some differences do exist. In particular, while the coefficients on female remain approximately the same (in terms of significance and sign), the coefficients on the interaction term with children do not. Instead, we find that adding children does not differentially affect the work status by gender of those with formal IT training. However, it does differentially affect the workforce status of women who had been working in IT who were not formally trained in IT (Table 6). We do not know the root cause of this differential. A possibility is that the "trained" group of women is more committed to the workforce than those who are not trained. The results also suggest some variation by minority status when the sample is split between the trained and the non-trained, but care must be taken given the small number of African Americans in the study which in and of itself may explain the change in significance.

Several other interesting results emerge when the sample is split between those with formal training and those without formal IT training For example, we now see that the variable indicating whether the individual were self-employed has a nuanced role depending upon training status. IT workers in 1993 who were formally trained in IT and who were self-employed in 1993 are significantly less likely than those who were not self employed to leave IT for another occupation. However, IT workers in 1993 who were not formally trained in IT and who were self-employed in 1993 are not less likely to leave IT for another occupation. They are, however, significantly more likely to leave IT for the "no work" category.

It is also worth noting that the change in visa status variable "gotperm" is only significant for those 1993 IT workers who did not have formal training. Specifically, it is those who

became permanent residents between 1993 and 1999 who were not formally trained in IT who are significantly more likely to leave IT for a non-IT occupation after they became permanent citizens; not those with formal training. This suggests that the H-1B visa hypothesis noted above applies to those without formal training, not those with formal training. Perhaps the former group is less committed to the occupation than the latter group and sees IT work primarily as a means to becoming a permanent resident.

**Table 5**. The odds of different employment outcomes in 1999 for those with formal training in IT and employed in IT occupations in 1993.

| | WORK IT vs. WORK NOT-IT | | WORK IT vs. NO WORK | | WORK NOT-IT vs. NO WORK | |
|---|---|---|---|---|---|---|
| Variables | Coeff. | Odds ratio | Coeff. | Odds ratio | Coeff. | Odds ratio |
| intercept | 1.315 | 3.725 | -1.338 | 0.262 | -2.653 | 0.070 |
| itlatest93 | ***0.810*** | 2.248 | 0.387 | 1.473 | -0.423 | 0.655 |
| female | -0.030 | 0.970 | ***-1.307*** | 0.271 | ***-1.277*** | 0.279 |
| african american | -0.486 | 0.615 | -0.076 | 0.927 | 0.410 | 1.507 |
| asian | 0.137 | 1.147 | -0.127 | 0.881 | -0.264 | 0.770 |
| hispanothers | 0.312 | 1.366 | 1.423 | 4.150 | 1.111 | 3.037 |
| perm93 | 0.162 | 1.176 | 0.515 | 1.674 | 0.354 | 1.425 |
| temp93 | 0.061 | 1.063 | 0.304 | 1.355 | 0.244 | 1.276 |
| age93 | -0.035 | 0.966 | ***0.312*** | 1.366 | **0.348** | 1.416 |
| agesq93 | 0.001 | 1.001 | ***-0.005*** | 0.995 | ***-0.006*** | 0.994 |
| addit | 0.255 | 1.290 | ***6.441*** | 627.0 | *na* | |
| addnoit | ***-1.288*** | 0.276 | -0.807 | 0.446 | 0.480 | 1.616 |
| gotmarried | -0.053 | 0.948 | 0.235 | 1.265 | 0.288 | 1.334 |
| gotsingle | 0.397 | 1.487 | -0.296 | 0.744 | -0.693 | 0.500 |
| gotchidren06 | 0.059 | 1.061 | -0.154 | 0.857 | -0.214 | 0.807 |
| gotperm | -0.093 | 0.911 | -0.449 | 0.638 | -0.356 | 0.700 |
| selfemp | ***-0.905*** | 0.405 | -0.616 | 0.540 | 0.287 | 1.332 |
| fgotmarried | -0.335 | 0.715 | -0.306 | 0.736 | 0.029 | 1.029 |
| fgotsingle | 0.327 | 1.387 | 0.702 | 2.018 | 0.376 | 1.456 |
| fgotchildreno6 | 0.054 | 1.055 | -0.728 | 0.483 | -0.782 | 0.457 |
| n | 2110 | | | | | |
| -2 log L | ***2311.8*** | | | | | |

Notes: Estimates indicated by ***XXX*** are statistically significant at the .01 level, by **XXX** at the .05 level, and XXX at the .10 level. *na* means that the parameter could not be estimated because of the lack of variation within the category. Variables indicated by "f" times the variable name denotes an interaction with the female dummy variable.

**Table 6**. The odds of different employment outcomes in 1999 for those without formal training in IT and employed in IT occupations in 1993.

| | WORK IT vs. WORK NOT-IT | | WORK IT vs. NO WORK | | WORK NOT-IT vs. NO WORK | |
|---|---|---|---|---|---|---|
| Variables | Coeff. | Odds ratio | Coeff. | Odds ratio | Coeff. | Odds ratio |
| intercept | **-2.030** | 0.131 | ***-7.196*** | 0.001 | ***-5.166*** | 0.006 |
| female | -0.082 | 0.921 | ***-0.577*** | 0.562 | **-0.495** | 0.610 |
| african american | -0.278 | 0.757 | 0.908 | 2.479 | 1.186 | 3.274 |
| asian | 0.301 | 1.351 | -0.040 | 0.961 | -0.341 | 0.711 |
| hispanothers | -0.049 | 0.952 | -0.243 | 0.784 | -0.194 | 0.824 |
| perm93 | -0.023 | 0.977 | -0.057 | 0.945 | -0.035 | 0.966 |
| temp93 | 0.495 | 1.640 | 1.031 | 2.804 | 0.536 | 1.709 |
| age93 | **0.117** | 1.124 | ***0.565*** | 1.759 | ***0.448*** | 1.565 |
| agesq93 | -0.001 | 0.999 | ***-0.008*** | 0.992 | ***-0.007*** | 0.993 |
| addit | 0.626 | 1.870 | *na* | | *na* | |
| addnoit | ***-0.967*** | 0.380 | 0.167 | 1.182 | 1.134 | 3.108 |
| gotmarried | 0.181 | 1.198 | 0.119 | 1.126 | -0.062 | 0.940 |
| gotsingle | -0.099 | 0.906 | -0.308 | 0.735 | -0.209 | 0.811 |
| gotchildren06 | -0.055 | 0.946 | 0.116 | 1.123 | 0.171 | 1.186 |
| gotperm | **-0.671** | 0.511 | -0.173 | 0.841 | 0.498 | 1.645 |
| selfemp | -0.223 | 0.800 | ***-0.665*** | 0.514 | -0.441 | 0.643 |
| othersci | ***0.634*** | 1.885 | ***0.601*** | 1.824 | -0.033 | 0.968 |
| othereng | 0.224 | 1.251 | ***0.729*** | 2.073 | 0.505 | 1.657 |
| bus | 0.019 | 1.019 | 0.278 | 1.320 | 0.259 | 1.296 |
| socsci | 0.136 | 1.146 | 0.366 | 1.442 | 0.230 | 1.259 |
| fgotmarried | -0.211 | 0.810 | -0.234 | 0.791 | -0.024 | 0.976 |
| fgotsingle | 0.614 | 1.848 | 1.201 | 3.323 | 0.586 | 1.797 |
| fgotchildreno6 | 0.392 | 1.480 | ***-1.468*** | 0.230 | ***-1.860*** | 0.156 |
| n | 3098 | | | | | |
| -2 log L | ***4507.8*** | | | | | |

Notes: Estimates indicated by ***XXX*** are statistically significant at the .01 level, by **XXX** at the .05 level, and XXX at the .10 level. *na* means that the parameter could not be estimated because of the lack of variation within the category. Variables indicated by "f" times the variable name denotes an interaction with the female dummy variable.

## V. Conclusions.

This study examines factors related to whether an individual working in IT is retained in an IT occupation. We are particularly interested in whether and why retention varies by gender and minority status. For purposes of analysis we use the SESTAT data, focusing on those who

are in the sample in both 1993 and 1999. Our empirical results are conditional on the individual working in IT in 1993.

We find that women and African Americans are less likely to be retained in an IT occupation than are men and whites. Moreover, this result is sustained when we estimate a logit model, holding other variables constant. This means that we are not able to "explain away" why women and African Americans are less likely to be retained than men and whites. We find no evidence that Hispanics and other underrepresented minorities are less likely to be retained than whites.

We find that differences exist between women and African Americans that manifest themselves when the data are analyzed using a multinomial logit model. Indeed, we find that the lower retention rate of women compared to men is due to the fact that women are more likely not to work. It is not because women leave IT for non-IT jobs. By way of contrast, the lower retention rate of African Americans is due to their leaving IT for other occupations. It is not because African Americans are observed not working in 1999.

We also find that the retention of women is significantly related to whether they begin parenting a young child during the period while the retention of men is not significantly related to the arrival of young children. Interestingly enough, it is the women who lack formal training in IT who are affected by the presence of young children, not women who are trained in IT.

Several other results are also worth noting. For example, we find no evidence that being self-employed enhances the probability of retention. Indeed, among those with IT training, it reduces it. Our results also suggest that temporary residents who lack formal training in IT and become permanent residents during the period are less likely to be retained than natives. This may reflect their freedom from H-1B visa status.

We conclude that policies directed towards retention will have differential outcomes depending upon the group in question. For example, with regards to retention, our results suggest that women, especially those without formal IT training, would be likely to respond to initiatives that provide on-site child care, thus making it less likely that they leave the labor force. By way of contrast, our results suggest that African Americans would be more likely to be responsive to initiatives that make IT occupations more appealing relative to non-IT jobs. Such initiatives could include mentoring and continuing education programs.

We must emphasize that the analysis is limited in several respects. First, it only examines patterns of retention for individuals who qualify to be included in the SESTAT data. It also has not identified the extent to which the results reflect discrimination. A case in point is that we cannot infer whether the finding that African Americans are more likely to work in non-IT occupations than are whites is due to discrimination. Another limitation is that the small number of underrepresented minorities in the sample means that we cannot simultaneously analyze how retention varies by minority status and gender. The results are also limited to the extent that they may be clouded by the strong labor market that existed for IT workers in the late 1990s, partly because of issues related to Y2K. In future work, we hope to examine how the labor market influences patterns of retention and recruitment by studying the relationship between retention and recruitment and local labor market conditions.

## Appendix: Glossary of Variable Names

| Variable | Definition |
|---|---|
| addit | =1, if additional education in IT from 1993-99; 0 otherwise. |
| addnoit | =1, if additional education in non-IT from 1993-99; 0 otherwise. |
| age93 | =age in 1993. |
| agesq93 | =age in 1993 squared. |
| african american | =1, if race/ethnicity was African American; 0 otherwise |
| asian | =1, if race/ethnicity was Asian; 0 otherwise. |
| bus | =1, if business was field of highest degree earned; 0 otherwise |
| female | =1, if female; 0 otherwise. |
| gotchildren06 | =1, if added children under the age of six from 1993-99; 0 otherwise. |
| gotmarried | =1, if became married from 1993-99; 0 otherwise. |
| gotperm | =1, if received permanent residency status from 1993-99; 0 otherwise. |
| gotsingle | =1, if became single from 1993-99; 0 otherwise. |
| hispanothers | =1, if race/ethnicity was Hispanic, Native American or other; 0 otherwise. |
| itlatest93 | =1, if the latest degree earned by 1993 was in an IT field; 0 otherwise. |
| No Work | =1, if not working in 1999; 0 otherwise. |
| othereng | =1, if not IT-trained and engineering was field of highest degree earned; 0 otherwise. |
| othersci | =1, if not-IT trained and physical or natural sciences was field of highest degree earned; 0 otherwise. |
| perm93 | =1, if non-U.S. citizen but permanent resident in 1993; 0 otherwise. |
| selfemp | =1, if self-employed in 1993; 0 otherwise. |
| socsci | =1, social sciences was the field of highest degree earned, 0 otherwise. |
| temp93 | =1, if non-U.S. citizen and temporary resident in 1993; 0 otherwise. |
| Work It | =1, if working in IT occupation in 99; 0 otherwise. |
| Work Not-IT | =1, if working in non-IT occupations in 99; 0 otherwise. |

Notes: All variables were derived from SESTAT. A site license for the individual-level data was obtained from NSF.

**References**


Alberts, D. S., D. S. Papp, eds. 1997. The Information Age: An Anthology of its Impact and Consequences. (www.ndu.edu/inss/books/Anthology//part1.html).

Allison, Paul D. 1999 Logistic Regression Using SAS System: Theory and Application. Cary, NC: Sas Institute Inc.

Barnow, Burt S., John Trutko, and Robert Lerman, 1998. "Skill Mismatches and Worker Shortages: The Problem and Appropriate Responses," Draft Final Report, The Urban Institute.

Bentsen, Cheryl. September 1, 2000. "Why Women Hate I.T." CIO Magaize, Sept. 1.

Boskin, M. J. 1974. "A Conditional Logit Model of Occupational Choice." Journal of Political Economy 82: 389-398.

Brown University, 1996. "Achieving Gender Equity in Science Classrooms: A Guide for Faculty." Compiled by Women Science Students and Science Faculty and Staff at NECUSE Colleges, Published by the Office of the Dean of the College.

Camp, T. August 2001. "Women in Computer Sciences: Reversing the Trend," Syllabus, pp. 24-28.

Carver, Doris L. 2000. "Research Foundations for Improving the Representation of Women in the Information Technology Workforce," Virtual Workshop Report, (www.cise.nsfgov/itwomen/itwomen_final_report.pdf).

Castells, M. 1996. The Rise of the Network Society. Cambridge, MA: Basil Blackwell.

Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development. 2000. Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering and Technology. (www.nsf.gov/od/cawmset/report/cawmset_report.pdf).

Congressional Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development, 2000. "Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering and Technology.

Department of Commerce, Office of Technology Policy. 1997. America's New Deficit: The Shortage of Information Technology Workers. (www.ta.doc.gov.).

Ellis, Mark and John Odland. 2001. "Intermetropolitan Variation in the Labour Force Paritication Rates of White and Black Men in the United States," Urban Studies, 38(13):2327-2348.

Ellis, Richard and B. Lindsay Lowell, 1999. IT Workforce Data Project, United Engineering Foundation. (www.uefoundation.org/itworkfp.html).

Freeman, C., L. Soete, and U. Efendioglu. 1995. "Diffusion and the Employment Effects of Infromation and Communication Technology." International Labor Review 134(4-5): 587-603.

Freeman, Peter and William Aspray. 1999. The Supply of Information Technology Workers in the United States. Washington, D.C: Computing Research Association.

Fullerton, Howard. 1999. "Labor Force Participation: 75 Years of Change, 1950-98 and 1998-2025," Monthly Labor Review, 22(12):3-12.

Garcia, Oscar N. and Roscoe Giles. 2000. Research Foundations on Successful Participation on Underrepresented Minorities in Information Technology, Final Report from a Cyberconference, (www.cise.nsf.gov/itminorities/itminorities_final_report.pdf).

Johnson, Susan Tinch and Jack A. Bobo. 1998. Software Workers for the New Millennium: Global Competitiveness Hangs in the Balance. Arlington, VA: National Software Alliance, (www.software_alliance.org).

Judge, George G., William E. Griffiths, R. Carter Hill, and Tsoung-Chao Lee. 1980. The Theory and Practice of Econometrics. New York: John Wiley and Sons.

Kranzberg, M. 1989. "IT as Revolution." In T. Forester, ed., Computers in the Human Context: Information Technology, Productivity, and People. Cambridge, MA: MIT Press.

Information Technology Association of American. Undated. "Building the 21st Century Information Technology Workforce: Underrepresented Groups in the Information Technology Workforce."

Lombard, Karen. 1999. "Women's Rising Market Opportunities and Increased Labor Force Participation," Economic Inquiry, 37 (April):195-212.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

Margolis, Jane and Allan Fisher, 2002. Unlocking the Clubhouse: Women in Computing. Cambridge, MA: MIT Press.

Matloff, Norman. 1988. "Debunking the Myth of a Desperate Software Labor Shortage." Testimony to the U.S. House Judiciary Subcommittee on Immigration, updated April 8, 2000. (ftp://heather.cs.ucdavis.edu/svreport.html).

National Resaerch Council. 2001. Building a Workforce for the Information Economy. Washington, DC: National Academy Press.

National Science Foundation, Division of Science Resource Studies. 1999. Retention of the Best Science and Engineering Graduates in Science and Engineering. Arlington, VA. [99-321].

_____. 1999. SESTAT: A Tool for Studying Scientists and Engineers in the United States. Authors, Nirmala Kannankutty and R. Keith Wilkinson. Arlington, VA. [NSF 99-337].

National Science Foundation. 2000. Women, Minorities, and Persons With Disabilities in Science and Engineering. Arlington, VA. [NSF 00-327].

Parsons, Donald. 1980. "Racial Trends in Male Labor Force Participation," American Economic Review, 70(5):911-920.

Office of Technology Policy, U.S. Department of Commerce. 1999. America's New Deficit: The Shortage of Information Technology Workers. (www.ta.doc.gov)

President's Information Technology Advisory Committee (PITAC). 1999. Report to the President, "Information Technology Research: Investing in Our Future." (http://www.ccic.gov/ac/report/).

Preston, Anne. 2004. Leaving Science: Occupational Exit from Scientific Careers, New York: Russell Sage Foundation.

Schmidt, P., and R. P. Strauss. 1975. "The Predictions of Occupation using Multiple Logit Models." International Economic Review 16(2): 204-212.

Scientific American, April 27, 1998. "Dropping Out: The 'Science Pipeline' Leaks Women at Every Step of Their Training. What are the causes?"

U.S. Bureau of Labor Statistics, Current Population Survey, http://data.bls.gov/labjava/outside.jsp?survey=ln.

Ward, Patricia A., Peter F. Orazem and Steffen W. Schmidt. 1992. "Women in Elite Pools and Elite Positions." Social Science Quarterly 73(1): 31-45.