The Relationship between Technology Revolution and College Graduates in the Field of Computing

Jaclyn Hawtin
Arizona State University, U.S.
May 2011

As humanity moves forward into the digital age of computing, there will be a great shift in the infrastructure of employment and education. In order for today's students to keep up with the demands and needs of our quickly evolving work force they must begin to start thinking about their education from a strategic and applicable perspective. The technology sector is one of the fastest growing industries and was projected to grow by 30% from 2004 to 2012 in the US (Horrigan, 2004). There has been much debate about whether or not the US population has been keeping up its required levels of education in the computing industry as a means to satisfy the growing demand. Many fear that the US may end up in the impervious position of needing to outsource a large majority of technology jobs as the US just does not have the workforce needed to support the growing demand. In this study we do an analysis of the growth rates of college graduates in the field of computing in the US from the years 1998 to 2008 to assess whether or not the percentage of US graduates in computing are increasing as time moves forward in contrast to other disciplines. Is there a noticeable difference in the percent of US students graduating with computing degrees between the years 1998 to 2008 in contrast to total graduation averages in all other fields? And does this data seem to be reflective of the general growth rates of the technology sector in the US from the years 1998 to 2008?

Literature Review

There has been a longstanding concern in the United States regarding the education of American citizens in the mathematics, science and engineering fields. We have noticed a large deviation of test averages in these areas in US public education systems when compared with many other countries. The National Center for Education Statistics (IES) did a cross comparative analysis in the year 2006 test scores among 15 year old students in the OCED countries and others who chose to join with 58 countries in total. In the mathematics literacy category the US ranked 36 with a score of 474 when the average was 498. In the category of science literacy the US ranked 30 with a score of 489 when the average was 500 (National Center for Education Statistics, 2006). It's difficult to comprehend why the world's largest super power could be lagging so far behind in levels of public educational competency in both science and math. Unfortunately the data for reading literacy levels was not available for the US in this analysis. But for the considerations of this paper the values for math and science provide a benchmark for historical US policy and intentions in the realm of education that is a necessary precursor for the growth of a population of successful computing professionals in the United States of America.

There are many in industry who argue that there is a shortage of qualified workers available in the US in the technology sector. Although critics and mainly academics have expressed skepticism about this shortage and claim that is greatly exaggerated. They claim that the idea of a so called shortage has been carefully crafted to keep labor costs low. Roberts makes three claims about the reasons there is such a disagreement about whether or not there is an actual labor shortage. The first claim states that most analysis has been done from a analysis on the whole perspective which does not take into account the dynamics of particular specialty areas where the shortages generally exist. The second claim recognizes the general view that individual workers in the software development industry and interchangeable which is generally a fallacy. The third claim states that industry is much more concerned with hiring individuals that are highly productive as opposed to reducing labor costs. An interesting thing about the technology industry is the variability in productivity of individual workers. "Programmers with the same level of experience exhibited variations of more than 20 to 1 in the time required to solve particular programming problems. Beyond this high variability in coding time, the study revealed that individual programmers showed highly correlated differences in other metrics that contribute to overall productivity, in the sense that the best programmers were not only able to complete the problem in less time, but in so doing typically produced programs that had fewer errors and were more efficient in both running time and utilization of memory (Roberts, 2000, p.3)." Because of these variations it is impossible to truly understand the dynamics of the labor market without factoring in this variability. Another interesting occurrence in individuals in the software industry is that talent does not necessarily result from a traditional education. There are many accounts of individuals that are self taught and have superseded general expectations, as in the case of Bill Gates. Even when cross comparing the abilities of those who did follow the route of attaining a traditional college education one can see that the variance of productivity in comparison to other industries is off the charts. One begins to question where this variation is coming from. Although today as systems are becoming more sophisticated the requirement for success in industry is almost always the traditional route (Roberts, 2000).

Because there is such fierce competition in industry for highly qualified computing professionals the incentive for those professionals to be in industry is much greater than it is for them to be in education. The bottom line is that they will make more money in industry. So the dilemma today is that many schools are unable to maintain a competitive edge because they are unable to recruit and maintain a workforce that can provide academic curriculum rigorous enough to produce the type of talent that industry is looking for. So the so called shortage that we see is not necessarily in the number of possible workers, it is more about the skill level of these qualified workers (Roberts, 2000).

IT can be seen as a driving force for the acceleration of US productivity. Stiroh found that, "particular, industries that made the largest investments in computer hardware, software, and telecommunication equipment in the 1980's and early 1990's show larger productivity gains after 1995." In his decomposition of aggregate productivity analysis he found that industries that were IT using or IT producing had a big impact on the US productivity revival which he thinks is attributed to their involvement with IT in general. As a cross comparison industries that were generally isolated from the IT revolution had a negative impact on the US productivity revival. Stiroh concludes that, "the strong and robust correlation between IT intensity and the subsequent productivity acceleration, however, implies that there may be a deeper relationship between IT investment and productivity growth." This analysis supports the thesis that in order for the US to be successful careful attention must be given to producing highly qualified professionals in the computing field.

In order to combat the lack of qualified professionals in the field of computing in the US two things must be done. The standards for education in math and science in the early years of American citizens must be increased, as well as those at the university level. Secondly we must find a way to inspire US students to participate in the field of computing, which will probably come more naturally once the foundational skill set has been established in their primary education.

As with any new field computing does not have children shouting I want to be a programmer when I grow up. This field is complex and difficult to understand, and is generally not understood by the public for the most part. In order to counter this trend extra attention must be spent in colleges and universities across the nation in making sure they are doing there part to recruit students into this field. Akbulut laid out 4 a model that describes the major variables included in the student decision making process. See below:

- **Self-Efficacy** - A student's judgment of his or her capability to perform effectively as a computing major.
- **Outcome Expectations** - A student's judgment regarding the likelihood that valued rewards will occur as a result of pursuing a computing major.
- **Interest** - An emotion that arouses attention to, curiosity about, and concern with a computing major.
- **Major Choice Goals** - A student's aspirations to choose a computing major.

US educational institutions are under pressure to produce qualified professionals in the field of computing. By utilizing the above factors that drive student decision making these academic institutions will have better success in converting students into computing professionals and secondly will be strengthening the US in ways that we now may not even be able to conceive of (Akbulut, 2007).

As we can now see that it is imperative for the US to continually produce a population of highly qualified computing students to support industry. But the question still stands, has the US been keeping up with this need.

Research Design & Methodology

In this analysis we collected data for two general areas. We started out by extracting a data set that included statistics on general graduation rates in the US from the years 1998 to 2008. The second step that we took was to focus down on the graduation rates for students in the US specifically in the area of computing, also from the years 1998 to 2008. At that point we had enough information to do an analysis on the trends in growth rates of graduates in the US as whole and as a comparison with the single field of computing. But as our research question states we wanted to know if there was a correlation between the growth rates in students graduating in the field of computing when compared with the growth rate of the US technology sector in general. The next step was to seek out information on the US technology sector. As there were a lack of direct statistics available as a whole for the US technology sector a number of variable indicators were extracted instead, and an index was created. Once the index had been established we could then correlate the dependent variable US computing student graduation rates with that of the growth in the technology sector in general (independent variable). Time here is also an independent variable.

The data used for this project was secondary; it was collected from two different online resources. The first set corresponding with US student graduate information was accessed from the UNESCO-OECD-Eurostat (OECD) database online. This data collection was on education statistics, and was compiled on the basis of national administrative sources, reported by Ministries of Education or National Statistical Offices. The units measured were number of graduates, variables collected included; distribution of graduates by country, year, level of education, program destination, program duration, program orientation, field of education and gender, the periodicity was yearly, and the reference period was the calendar year.

The graduates included in this data set encompass those who have graduated during the reference year of data. Any student who does no complete their final year of school but later gains an equivalency in the form of a test based on education from outside of this system are not considered 'graduates' in this data set. A successful completion is determined by the requirements set forth by the United States; these requirements may include a final examination series or total number of experiential hours accomplished for a specific discipline.

Each qualifying variable for 'graduate' also included a number of sub-defining categories. As level of education can be considered a deleterious title it must be explained. The statistical title for level of education utilized in the UOE database read *905160: Tertiary-type A and advanced research programs*, see definition below:

> "Tertiary-type A programs (ISCED 5A) are largely theory-based and are designed to provide sufficient qualifications for entry to advanced research programs and professions with high skill requirements, such as medicine, dentistry or architecture. Tertiary-type A programs have a minimum cumulative theoretical duration (at tertiary level) of three years' full-time equivalent, although they typically last four or more years. These programs are not exclusively offered at universities. Conversely, not all programs nationally recognized as university programs fulfill the criteria to be classified as tertiary-type A. Tertiary-type A programs include second degree programs like the American Master. First and second programs are sub-classified by the cumulative duration of the programs, i.e., the total study time needed at the tertiary level to complete the degree. "
> **Source:** http://stats.oecd.org/glossary/detail.asp?ID=5440

The variable program destination includes; further education/ theoretically based programs, and further education/ practical/technical/occupationally specific programs. Program duration encompasses anything from three to five years, less than five years, five to six years, five years or more, and more than six years. The program orientation can include; vocational and technical programs, first degree/qualification, and second qualification degrees. This was not a gender specific study and did include all males and females, although not the topic of this study these data points are available for further analysis. The data comes from, The Organization for Economic Co-operation and Development (OECD), and was established on September 30th

1961. This organization consists of about 40 countries, all of them in cooperation of identifying economic problems and working together to solve them.

The second database of information accessed was drawn from the World Bank, this included the variables used to create an index for growth in the US technology sector between the years of 1998 and 2008. These variables Include; labor participation rate, total (% of total population ages 15+), high-technology exports (% of manufactured exports), high-technology exports (current US$), patent applications (nonresidents), patent applications (residents), research and development expenditure (% of GDP), researchers in R&D (per million people), scientific and technical journal articles, trademark applications (direct nonresident), trademark applications (direct resident), and trademark applications (total). Two variables that were included in this section on the World Bank website under labor statistics and technology were technicians in R&D (per million people) and trademark applications (aggregate direct). These variables contained no data for the specified requested data set, and their absence may contribute to the range of error in the following computations.

The population size that we were dealing with in these data sets was the entire population of students graduating in computing in the United States, this number varied by year, by 2008 there were 57,261 graduates in the US the field of computing, and there were 2,343,517 graduates in all fields total. The population statistics for the technology industry indicators variables included; numerical values, percentages, and dollars, these values were also representative of the entire population, no samples were used. These data sets had to further broken down in order to create an index. Instead of using the hard data, all of the information was converted into growth rate statistics between each year so that the units were the same across all sets. Both the data from UOE and the World Bank was converted into growth rate percentages for ease of use in the following statistical analysis models.

A number of descriptive statistics tables were utilized in this analysis including crosstabs and graphs. The main statistical method employed was correlation-regression analysis to test whether or not the number of students graduating in computational fields follows a similar growth pattern to growth rates in the technology sector in the United States. After laying out the data for the growth rates in the technology sector and the growth rates of computing graduates in a crosstab I suspected there was an association between these two data sets. So in the next step I went on to calculate the measures of association to test whether or not this assumption was correct.

Research Analysis

US Graduation Summary Statistics with Focus on Computing

Year	US Computing Graduates		US Graduates All Other Fields		Total US Graduates
1998	35,723	(2.08%)	1,681,163	(97.92%)	1,716,886
1999	38,956	(2.24%)	1,700,222	(97.76%)	1,739,178
2000	51,236	(2.82%)	1,768,559	(97.18%)	1,819,795
2001	58,760	(3.20%)	1,778,497	(96.80%)	1,837,257
2002					
2003	78,358	(3.94%)	1,909,434	(96.06%)	1,987,792
2004	80,540	(3.85%)	2,009,361	(96.15%)	2,089,901
2005	73,646	(3.42%)	2,080,156	(96.58%)	2,153,802
2006	65,951	(2.97%)	2,157,078	(97.03%)	2,223,029
2007	59,997	(2.63%)	2,219,382	(97.37%)	2,279,379
2008	57,261	(2.44%)	2,286,256	(97.56%)	2,343,517

Source: *Organization for Economic Co-operation and Development (OECD) online database.*

In the above table of summary statistics you can see the percent of graduates in the computing field from the years 1998 to 2008. Although the numbers do seem to go up and down throughout this time period there is an overall increase of graduates in the field of computing of .36% which accounts for an increase of 21,538 graduates since 1998. This increase is not directly correlated with the percentage as the total US population of graduates grew throughout this time period as well. Percent increase is a more appropriate measure when analyzing the growth of graduates in the computing field as a whole.

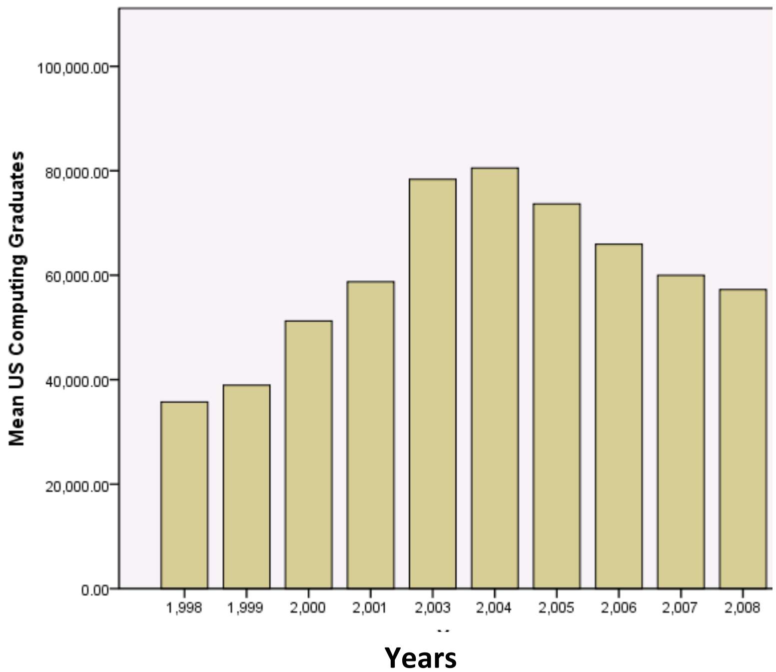
Cross-tabulation of % Growth in US Computing Graduates & Us Technology Sector

Years	% Growth Computing Graduates US	% Growth Technology Sector US
98-99	9.05%	2.69%
99-00	31.52%	8.28%
00-01	14.68%	-5.64%
01-02		-1.94%
02-03		0.99%
03-04	2.78%	3.11%
04-05	-8.56%	3.70%
05-06	-10.45%	4.97%
06-07	-9.03%	4.03%
07-08	-4.56%	-0.86%

Source: *Organization for Economic Co-operation and Development (OECD) online database & World Bank online database.*

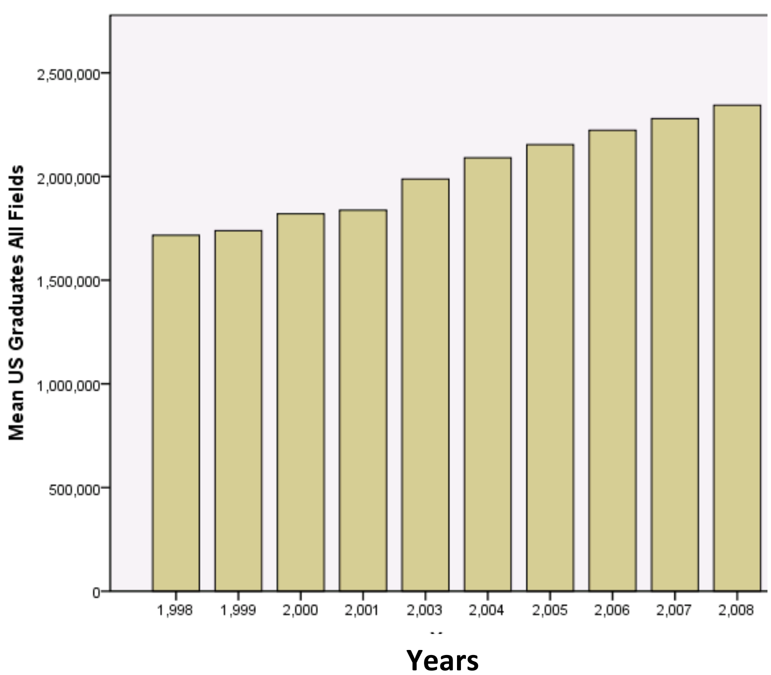
In the crosstabulation above one can see that there was a fluctuation in growth rates for both the technology sector as well as in computing graduates. The growth in the technology sector dips between the years of 2000 and 2002, whereas in computing graduates it dips in 2004-2008. This could potentially be a residual effect that the technology sector had on graduation rates later on, but further analysis would need to be done in order to make that conclusion. In the graphs below one can see a steady rise in the graduation rates of US students on the whole, but that dip in computing graduates is more apparent when looking at it in graphic format.

Graph 1: US Computing Graduates 10 Year Perspective (1998-2008)



Source: Organization for Economic Co-operation and Development (OECD) online database.

Graph 2: US Graduates All Fields 10 Year Perspective (1998-2008)



Source: Organization for Economic Co-operation and Development (OECD) online database.

Linear Regression Analysis

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	% Increase Technology Index[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: % Increase Computing Majors

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.072[a]	.005	-.161	15.777

a. Predictors: (Constant), % Increase Technology Index

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.804	1	7.804	.031	.865[a]
	Residual	1493.454	6	248.909		
	Total	1501.258	7			

a. Predictors: (Constant), % Increase Technology Index

b. Dependent Variable: % Increase Computing Majors

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.542	6.656		.382	.716
	% Increase Technology Index	.254	1.433	.072	.177	.865

a. Dependent Variable: % Increase Computing Majors

After running a statistical analysis between the values provided for percent increase in computing graduates in the US from the years 1998 and 2008 and the values for percent increase in the growth of the technology sector in the US from the years 1998 to 2000 we can conclude that there is very little to no association between the two. Pearson's r is only .072, and the coefficient of determination is a very small .005. Although, this test is not conclusive of a lack of association between these two data sets. There are many things that may have been overlooked. For instance the data set regarding growth in the technology sector may not have been representative, the variables included in this calculation were missing in two categories, and more research may have been needed to better represent what variables that were required to assess the US technology sector as a whole. Although we did not find a correlation between these two data sets some things can be recognized. US computing graduate numbers fluctuated a lot between the years 1998 and 2008, this movement could represent the settling of the industry as a whole as it finds a way to fit into the US economy. The graduation rates have also increased in the field of computing as compared with that of other fields from 1998 to 2008 suggesting that this is a field that is on the rise. Although the increase is less than one percent in context of all of the other competing fields this should still be recognized as an indicator that change that is occurring in US graduation rates. One factor that was not discussed here but should be recognized is that the graduation rates of US schools also include international students. This is something that should be explored further so that we can have a realistic idea of what percentage of those computing graduates are actually US citizens. As the statistics above showed that the US is lacking in educational levels at 15 years old, it would follow that it is very likely a large majority of computing majors in American schools are not actually American citizens. In conclusion we can say that the analysis above suggests that the percent increase in US computing graduates from 1998 to 2008 is not affected by the percent increase in growth rates of the US technology sector.

Bibliography:

1. Organisation for Economic Co-operation and Development (OECD). (2011). *Graduates by field of education* [Data file]. Retrieved from http://stats.oecd.org/Index.aspx
2. Horrigan, Michael W (2004) Employment Projections to 2012: concepts and context. *Monthly Labor Review*, 22, 20 pages.
3. Roberts, Eric (2000). Computing Education and the Information Technology Workforce. Retrieved from http://www.acm.org/education/.
4. Stiroh, Kevin J. (2002). Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say? *The American Economic Review*, Vol. 92, No. 5.
5. Akbulut, Asli Yagmur & Looney, Clayton Arlen (2007). Inspiring Students to Pursue Computing Degrees. *Communications of the ACM* October 2007/Vol. 50, No. 10
6. Biggers, Maureen & Brauer, Anne & Yilmaz, Tuba (????). Student Perceptions of Computer Science: A Retention Study Comparing Graduating Seniors vs. CS Leavers
7. National Center for Education Statistics. (2006). Average mathematics literacy, reading literacy, and science literacy scores of 15-year-old students, by sex and country [data file]. Retrieved from http://nces.ed.gov/programs/digest/d09/tables/dt09_402.asp
8. World Bank. (2011). Infrastructure & Technology. Retrieved from http://data.worldbank.org/

Appendices:

US Statistical Indicators for Technology Industry

Technology Indicators	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Labor participation rate, total (% of total population ages 15+)	66.7	66.7	66.8	66.4	66.1	65.7	65.4	65.4	65.6	65.4	65.4
High-technology exports (% of manufactured exports)	33.22433	34.2151	33.65616	32.58824	31.81089	30.80521	30.33195	29.92186	30.07804	28.53806	27.12218
High-technology exports (current US$)	1.72E+11	1.81E+11	1.97E+11	1.76E+11	1.62E+11	1.6E+11	1.77E+11	1.91E+11	2.19E+11	2.29E+11	2.31E+11
Patent applications, nonresidents	102246	116512	131100	148958	150200	153500	167407	182866	204182	214807	224733
Patent applications, residents	134733	149251	164795	177513	184245	188941	189536	207867	221784	241347	231588
Research and development expenditure (% of GDP)	2.610057	2.664309	2.745791	2.761431	2.659595	2.656179	2.580505	2.611706	2.650799	2.715062	2.824645
Researchers in R&D (per million people)		4431.605	4480.858	4535.156	4566.02	4817.842	4647.837	4584.387	4663.283		
Scientific and technical journal articles	190431.4	188004.1	192743	190592.6	190496.1	196431.6	202084.3	205516.3	209237.2	209694.7	
Trademark applications, direct nonresident	39182	31040	41244	34595	30944	28825	27801	28359	30255	33065	32105
Trademark applications, direct resident	207429	229721	251220	181713	181693	191902	213495	224269	233311	256429	246220
Trademark applications, total	246611	260761	292464	216308	212637	220965	248406	264510	277579	304129	294070
Technicians in R&D (per million people)											
Trademark applications, aggregate direct											

Source: World Bank

Note: I included the entire linear regression analysis in the paper, as I thought you would want to be able to see all of the numbers. Also because I was working with population numbers I was not able to do any of the hypothesis tests because I was not working with a sample. The t-tests from earlier on required means… which did not fit my data set either, so it seemed that the only plausible test that could be used was a regression correlation between the two data sets.