"The Effects of Specialization, Age, and Utilization of Traditional Tools on Perception of Google as an Effective Information Tool"

Michael J. Giarlo

17:610:511:01 – Professor Dan O'Connor, Ph.D.

March 28, 2005

INTRODUCTION

The impact of the Information Age upon our lives grows daily, as more and more of us become connected to the rest of the world via the Internet, an expansive, evolving network of devices which are, for the most part, designed to understand and convey only two concepts: 0 and 1. A vast amount of these 0s and 1s is stored on computers, in formats which we have designed to make sense of it all, and made accessible over the Internet, thanks in large part to the efforts of pioneers in the fields of computer science and library/information science (LIS): whereas the former provided the groundwork and the vision, the latter provided order and conceptual systems of information retrieval based partly upon tried-and-true methods which have been tested for years in libraries and information centers.

The nervous whispers about the chaotic nature of information on the Internet and, more importantly, about library users' increasing reliance upon the Internet rather than the library for the satisfaction of information needs have been heard in the hallowed halls of academic libraries and echoed in the scholarly LIS literature. Popular tools such as Google, the current undisputed champion of Internet search engines, have been demonized by

some library science academics, though whether it is done in spite of, or because of, their popularity is unclear. What has been made clear in much of the literature is that Google is juxtaposed with more traditional information retrieval tools, such as online public access catalogs (OPACs), leading the reader to wonder what may be said about the connections between such tools.

THE MODELS

The aim of this study is to determine the effects of age, specialization, and usage of traditional tools upon usage of Google, specifically within the campus library environment. It is expected that age may be an important factor as, generally speaking, digital technologies have been more widely adopted by the younger generations due to their comparatively earlier exposure to such technology. Taking into account much of the LIS literature, specialization in the LIS field is seen as another potential concept that may be relevant to attitudes about non-traditional information tools such as Google. Finally, the juxtaposition of Google with the OPAC, as evinced in much of the field's literature, is also considered a relationship worth scrutinizing.

The concepts of the theoretical model are operationalized as follows in the empirical model:

- Age – the number of years the user has been alive
- Specialization – whether or not a user has an academic degree within the LIS field
- Usage of Traditional Tools – the number of hours the user has utilized the campus library's OPAC during the most recent week
- Usage of Google – the numbers of hours the user has utilized the Google search engine during the most recent week

The empirical model may thus be stated as "the effects of number of years alive, attainment of an LIS degree, and the number of hours spent using the OPAC last week upon the number of hours spent using Google last week."

METHODOLOGY

The unit of analysis for this study is adult persons, specifically those within the campus library. In order to tease out any relationships between the usage of Google and the usage of OPACs, the environs of the campus

library would seem to be a suitable location from which to draw users, though any similar such information center should suffice. Due to the rather narrow scope applied to the unit of analysis, the population from which the sample for this study is drawn is described as "campus library users," where the term "users" may equally be applied to college students, faculty, library staff, and any of the general public who happen to be using the campus library on the day during which the study is conducted.

The instrument chosen to gather the raw data for this study is a survey consisting of four questions: "What is your age (in years)?", "Do you have a degree in Library/Information Science?", "How many hours have you spent using the library's online catalog this week?", and "How many hours have you spent using Google this week?" In order to strengthen the validity of the conclusions drawn herein, selection of users is done randomly. Specifically, the instrument is distributed every twenty minutes to a user exiting the campus library. Distribution occurs during five-hour windows on consecutive weekdays, one 9am-1pm window and one 1pm-6pm window, during the middle of the Spring semester, in order to represent both early morning users and late afternoon users and also to avoid the possibility of delivering results which may be skewed by gathering all data

on a particular day.

The variables in this study are age, specialization, usage of OPAC, and usage of Google. The first independent variable, age, coded as "IV1," is a continuous, ratio variable ranging from 27 to 58. Another independent variable, specialization in LIS, coded as "IV3", is a categorical, nominal variable measured by a value of "Yes" or "No" stating whether the subject has an LIS degree. The final independent variable, usage of OPAC, coded as "IV2," is a continuous, ratio variable representing how many hours the user has spent using the OPAC over the past week. The dependent variable, usage of Google, is also a continuous, ratio variable similarly representing how many hours the user has spent using Google over the past week, and is coded as "DV" in the raw data.

DATA ANALYSIS

Univariate Statistics and Displays

The indexes of central tendency for IV1, age, indicate that the average age of campus library users during the times the study was conducted was

Age	
Mean	42.07
Standard Error	1.32
Median	42
Mode	38
Standard Deviation	7.23
Sample Variance	52.20
Kurtosis	-0.15
Skewness	0.10
Range	31
Minimum	27
Maximum	58
Sum	1262
Count	30

around 40, with both the mean and the median at 42 and the mode at 38. The range of ages among respondents was bounded by 27 as the minimum and 58 as the maximum, while the variance of the sample was determined to be 52.2. The standard deviation for age, derived from the value of the variance, was 7 (rounded down for simplicity of calculation), and the distribution of age scores may be characterized in terms of standard deviation as follows:

73.3% of subjects between +1SD (49) and –1SD (35)

13.3% of subjects between –2SD (28) and –1SD (35)

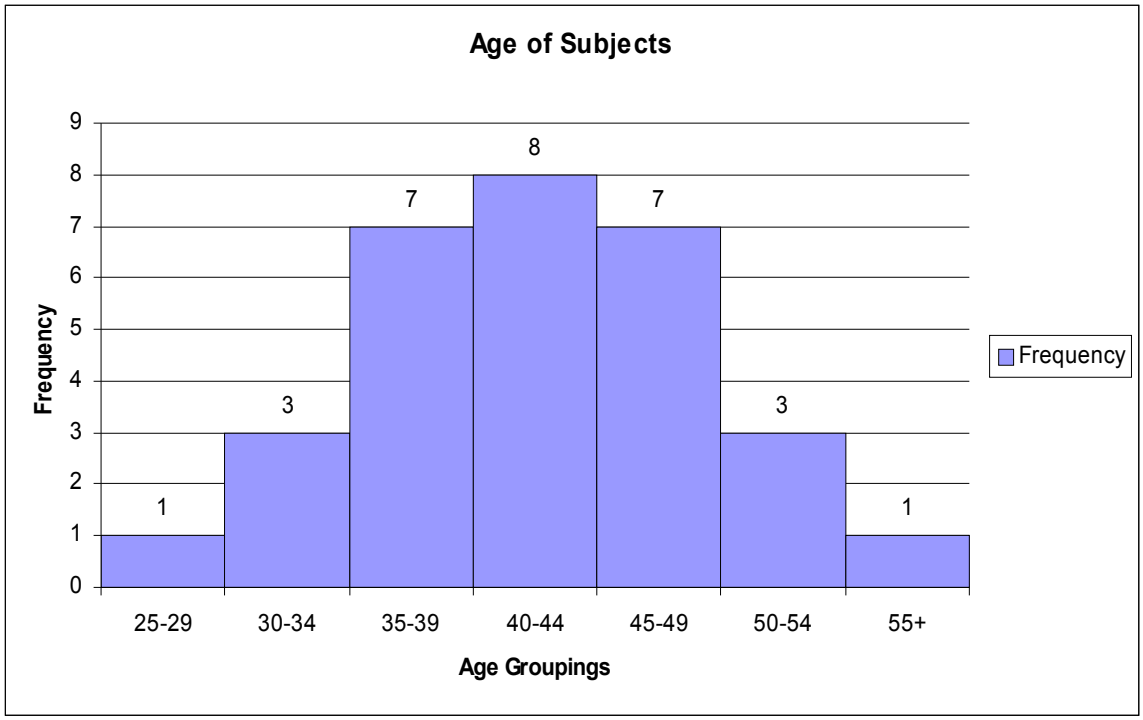
13.3% of subjects between +1SD (49) and +2SD (56)

3.3% of subjects greater than +2SD (56)

3.3% of subjects less than –2SD (28)

Compared to the normal frequency distribution, a higher percentage of scores fall into the range bounded by +1SD and –1SD by 5.3%. The differences between the sample and normal distributions in the other areas are less notable, with a deviation of only 0.2% between –2SD and –1SD and also between +1SD and +2SD and a deviation of 0.8% for scores greater than +2SD and less than –2SD. The “shape” of the distribution, then, is

slightly more peaked than a normal distribution as evidenced by the greater number of scores between +1SD and –1SD. The following figure is a frequency histogram of the age scores grouped into seven bins for ease of display.



The indexes of central tendency for IV2, usage of OPAC, indicate that the average number of hours spent using the OPAC during the past week was approximately 4, with both the mode and the median at 4 and the mean at 3.97. The range of hours among respondents was bounded by 0 as the minimum and 7 as the maximum, while the variance of the sample was

Hrs/Wk OPAC	
Mean	3.97
Standard Error	0.28
Median	4
Mode	4
Standard Deviation	1.54
Sample Variance	2.38
Kurtosis	0.21
Skewness	-0.30
Range	7
Minimum	0
Maximum	7
Sum	119
Count	30

determined to be 2.38. The standard deviation for OPAC usage, derived from the value of the variance, was 1.5 (rounded down for simplicity of calculation), and the distribution of scores may be characterized in terms of standard deviation as follows:

66.6% of subjects between +1SD (5.5) and –1SD (2.5)

13.3% of subjects between –2SD (1) and –1SD (2.5)

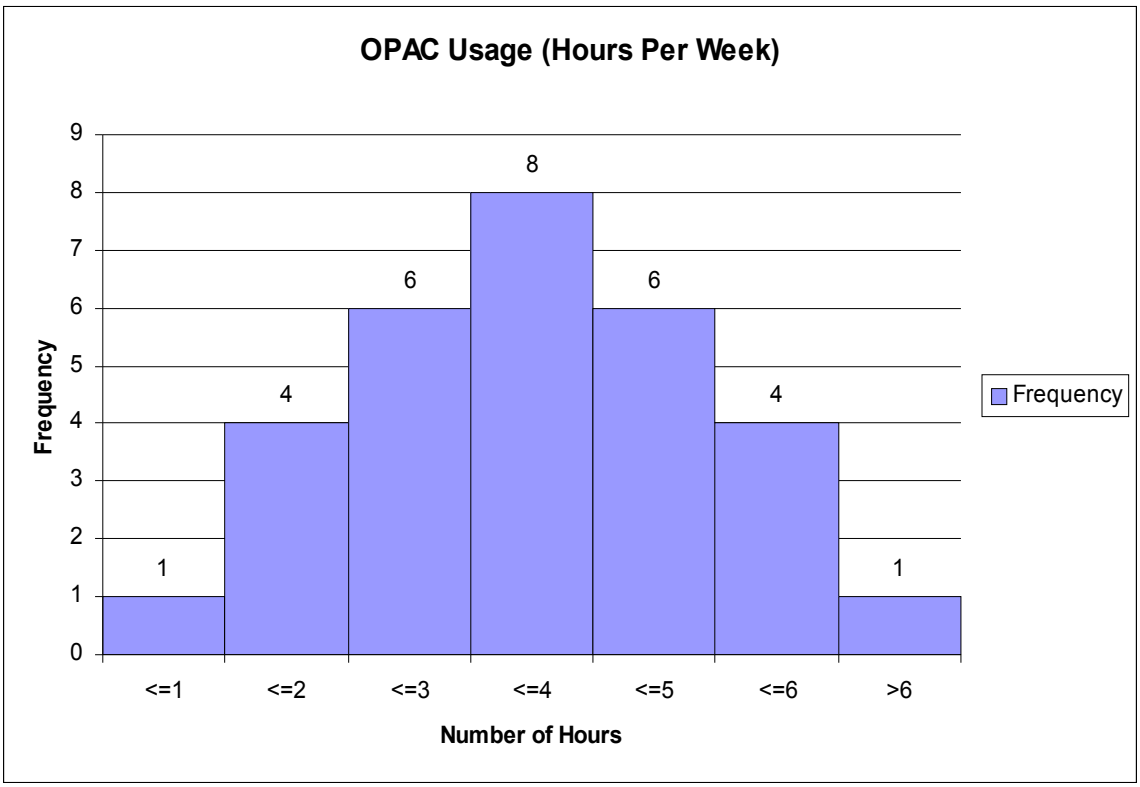
16.6% of subjects between +1SD (5.5) and +2SD (7)

0% of subjects greater than +2SD (7)

3.3% of subjects less than –2SD (1)

Compared to the normal frequency distribution, a nominally lower percentage of scores fall into the range bounded by +1SD and –1SD by 1.4%. The differences between –2SD and –1SD and less than –2SD are similarly nominal, at 0.2% and 0.8% respectively. The differences between the sample and normal distributions in the other areas are a bit more notable, with a deviation of 3.1% between +1SD and +2SD and a deviation of 2.5% for scores greater than +2SD. That is, there are no scores above two standard deviations from the mean, so the dispersion is slightly skewed towards the left. The "shape" of the distribution, overall, is about as peaked

as a normal distribution and is skewed marginally leftward. The following figure is a frequency histogram of the OPAC usage scores grouped into seven bins for ease of display.



The indexes of central tendency for DV, usage of Google, indicate that the average number of hours spent using Google during the past week was 4, with the mode, median, and mode calculated to be exactly 4. The range of hours among respondents was bounded by 1 as the minimum and 7 as the maximum, while the variance of the sample was determined to be 2.14. The standard deviation for Google usage, derived from the value of the variance,

Hrs/Wk Google	
Mean	4
Standard Error	0.27
Median	4
Mode	4
Standard Deviation	1.46
Sample Variance	2.14
Kurtosis	-0.53
Skewness	0.00
Range	6
Minimum	1
Maximum	7
Sum	120
Count	30

was 1.5 (rounded up for simplicity of calculation), and the distribution of scores may be characterized in terms of standard deviation as follows:

66.6% of subjects between +1SD (5.5) and –1SD (2.5)

16.6% of subjects between –2SD (1) and –1SD (2.5)

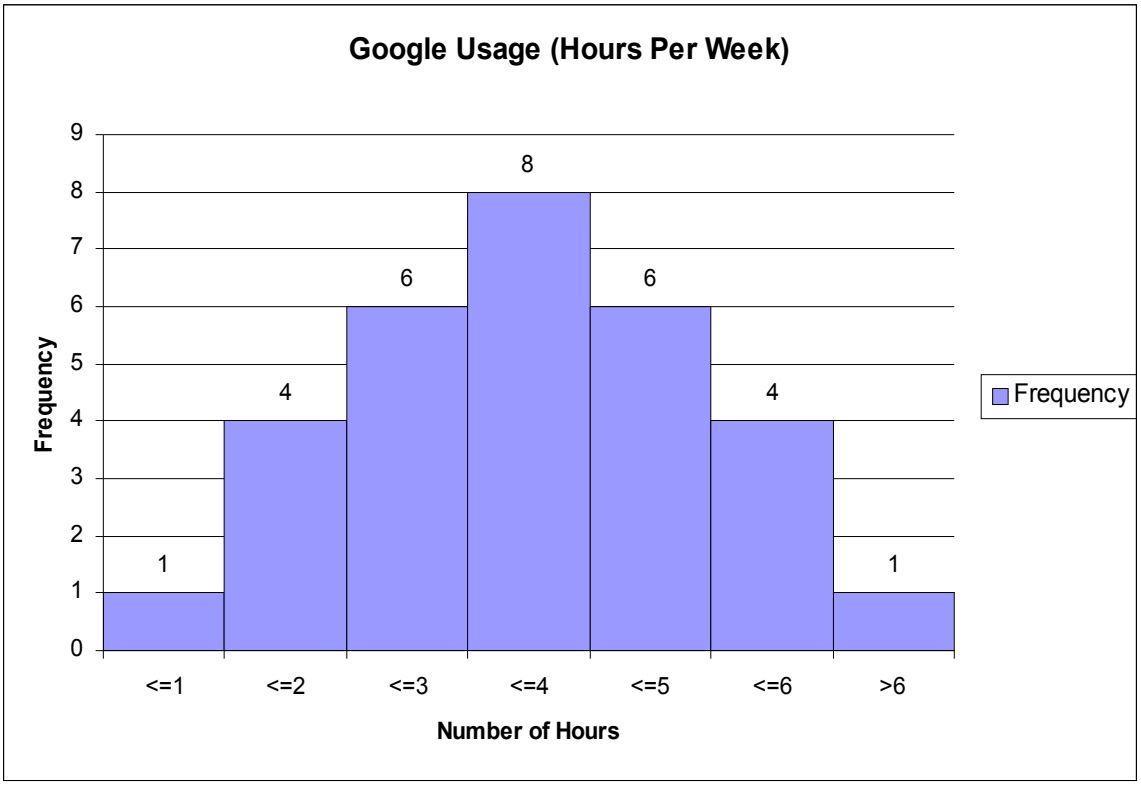
16.6% of subjects between +1SD (5.5) and +2SD (7)

0% of subjects greater than +2SD (7)

0% of subjects less than –2SD (1)

Compared to the normal frequency distribution, a nominally lower percentage of scores fall into the range bounded by +1SD and –1SD by 1.4%. The differences between the sample and normal distributions in the other areas are a bit more notable, with a deviation of 3.1% between +1SD and +2SD and –2SD and –1SD, and a deviation of 2.5% for scores greater than +2SD and less than –2SD. That is, there are no scores above or below two standard deviations from the mean, so the dispersion is slightly higher in the areas between one and two standard deviations from the mean, both on the positive and negative sides. The "shape" of the distribution, overall,

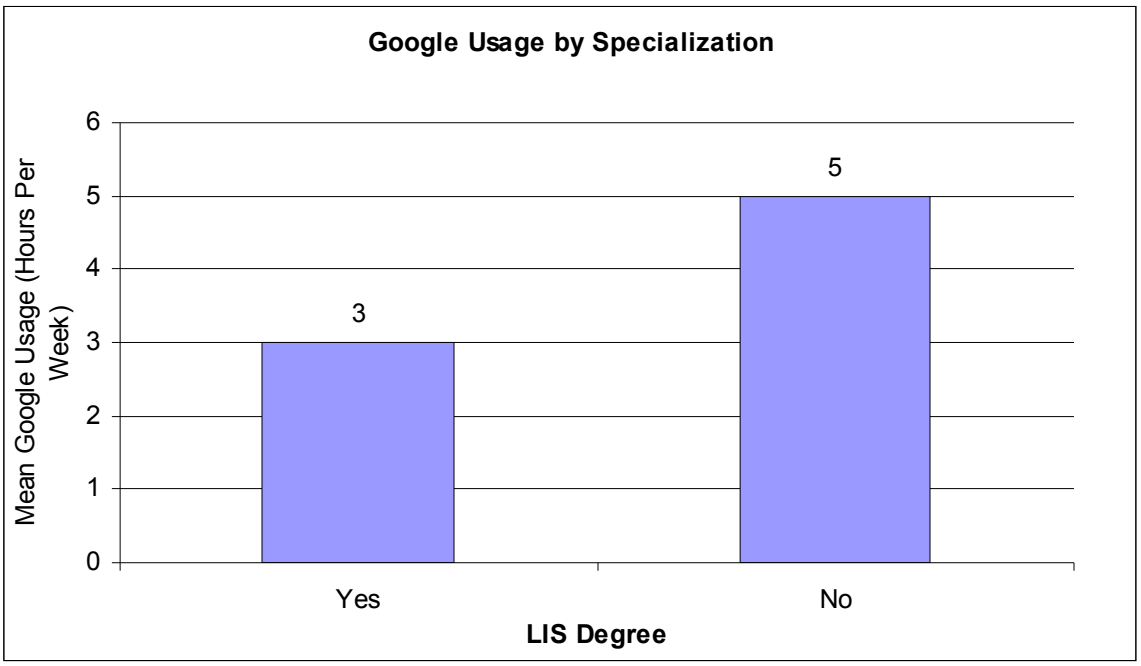
is about as peaked as a normal distribution but has boundaries that are quite a bit closer to the mean since no scores fall outside of two standard deviations from the mean. The following figure is a frequency histogram of the Google usage scores grouped into seven bins for ease of display.



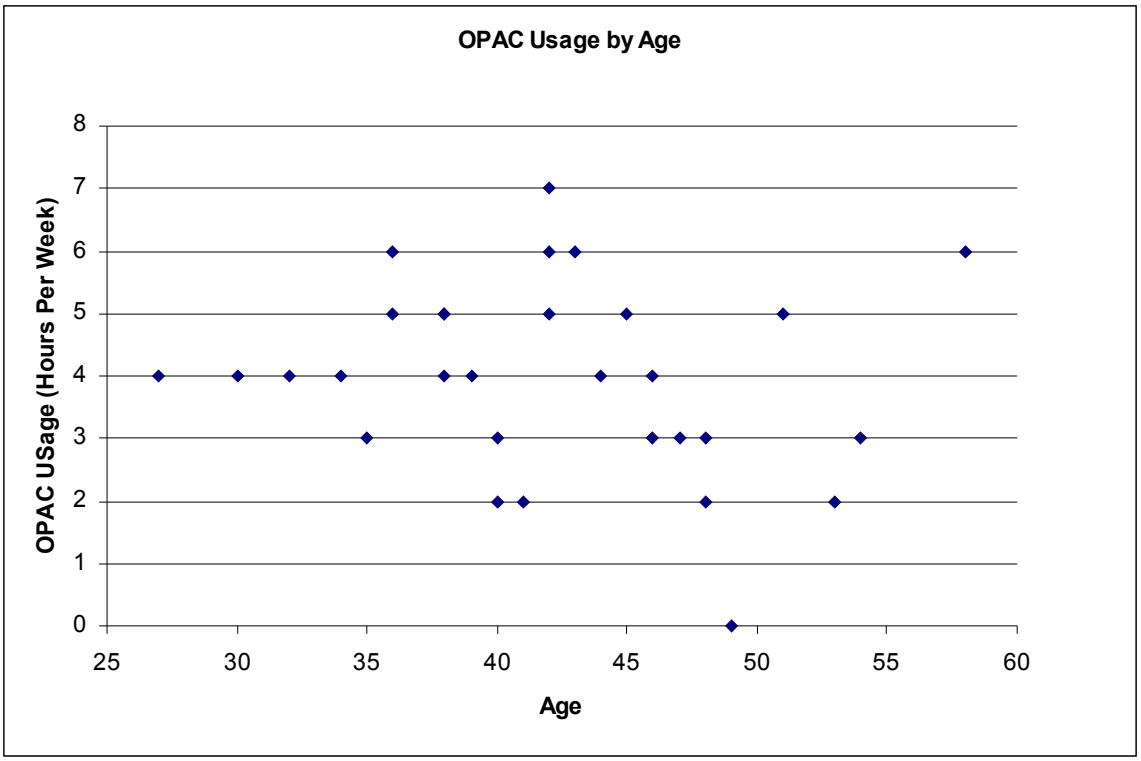
Bivariate Statistics and Displays

The bivariate statistics displays below include one bar chart showing the relationship between the DV, usage of Google, and IV3, Specialization, and three scatterplots showing the relationships between the DV and IV1, Age,

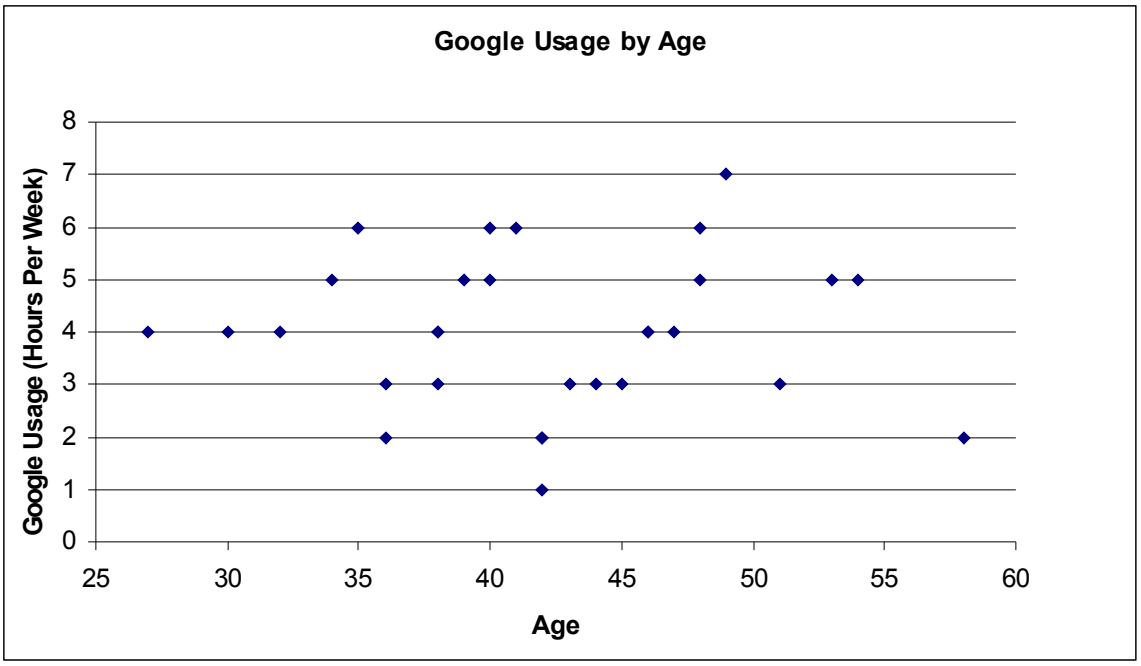
the DV and IV2, usage of OPAC, and IV2 and IV3.



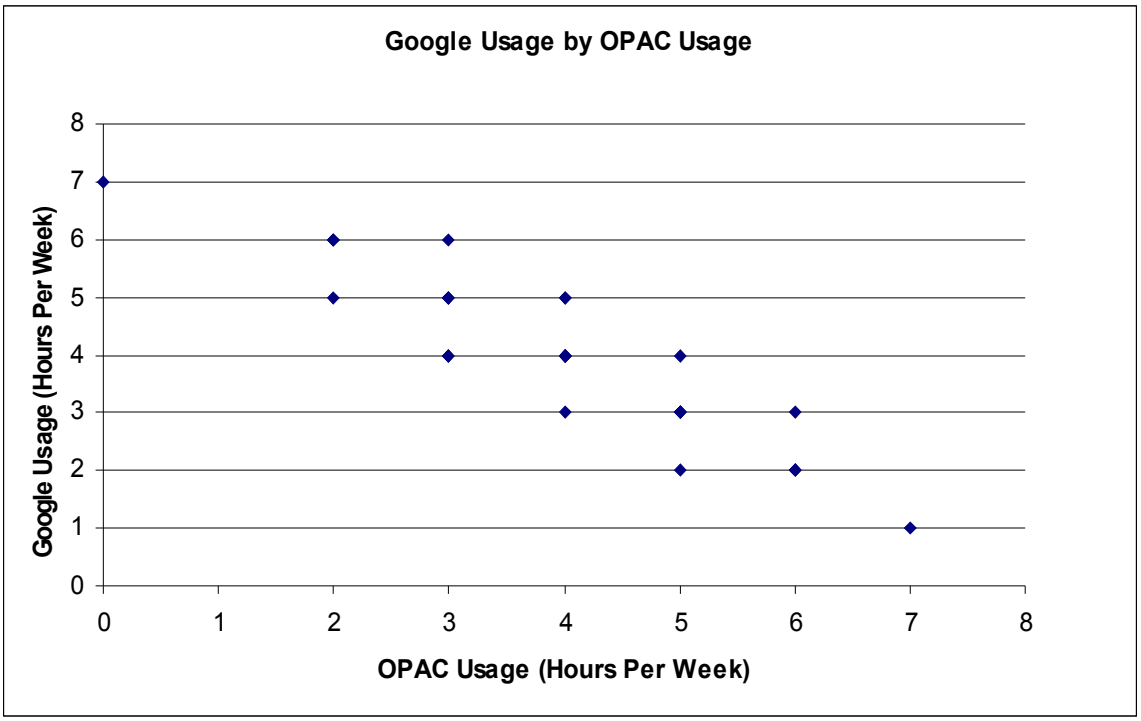
The bar chart figure above shows the relationship between the DV, usage of Google, and IV3, Specialization. Half of the subjects had received an LIS degree, and the other half had not. The mean value of DV for those with an LIS degree was 3 hours during the past week, compared with 5 hours for those without an LIS degree. This picture of the data indicates that those who specialize in LIS, within the particular campus library environment, use Google 40% less than those who do not specialize in LIS.



The scatterplot figure above shows the relationship between IV2, usage of OPAC, and IV1, Age. This picture of the data, with scores falling roughly throughout the entire chart, indicates that there is no clear relationship between age and OPAC usage.



The scatterplot figure above shows the relationship between DV, usage of Google, and IV1, Age. This picture of the data, with scores falling roughly throughout the entire chart, indicates that there is no clear relationship between age and Google usage.



The scatterplot figure above shows the relationship between DV, usage of Google, and IV2, usage of OPAC. This picture of the data, with scores generally trending downward, indicates that there is an inverse relationship between Google usage and OPAC usage. That is, as usage of OPAC goes up, usage of Google goes down.

CONCLUSION

Based upon the data collected in this study, it is concluded that some of the variables that have been tested exhibit a clear relationship while others do not. Specifically, age does not seem to have a relationship with either usage of Google or usage of the campus OPAC, counter to expectations before the study was conducted. On the other hand, specialization in LIS and usage of the campus OPAC seem to have a more or less strong relationship to usage of Google with a 40% drop in Google usage for those with an LIS degree and a clear negative trend for those who use the OPAC more frequently.

Improvements could be made upon this study and are suggested herein for those who wish to extend the investigation. Firstly, a much greater sample size would unquestionably yield more conclusive results. Secondly, the study would ideally be spread out over a longer period of time to account for temporal anomalies, for example, conducting a study during a week that coincides with an abnormally high number of bibliographic instruction courses which may skew the results towards artificially high OPAC usage. Thirdly, as every campus has its own culture and values, the study ought to be conducted at different colleges and universities, both geographically

different and “demographically” different, so to speak, such as a four-year, urban ivy and a two-year, suburban community college. Lastly, it is suggested that although age did not prove to be a significant variable in this study, it ought to be investigated more close in other studies. That the sample taken in this study ranges from 27 to 58 is seen as something of an anomaly; normally, the range would likely be spread out quite a bit more, and a larger sample size would bear this out in all likelihood. With a more normal distribution of age scores, this variable may yet prove to be significant.

RAW DATA

IV1	*IV2*	*IV3*	*DV*
Age	**Hrs/Wk OPAC**	**LIS Degree**	**Hrs/Wk Google**
27	4	N	4
30	4	N	4
32	4	N	4
34	4	N	5
35	3	N	6
36	5	N	3
36	6	Y	2
38	4	N	4
38	5	Y	3
38	5	Y	4
39	4	N	5
40	3	N	5
40	2	N	6
41	2	N	6
42	7	Y	1
42	6	Y	2
42	5	Y	2
43	6	Y	3
44	4	Y	3
45	5	Y	3
46	4	Y	4
46	3	Y	4
47	3	Y	4
48	3	N	5
48	2	N	6
49	0	N	7
51	5	Y	3
53	2	N	5
54	3	Y	5
58	6	Y	2

Age Bins	Frequency
25-29	1
30-34	3
35-39	7
40-44	8
45-49	7
50-54	3
55+	1

Hrs/Wk OPAC Bins	Frequency
<=1	1
<=2	4
<=3	6
<=4	8
<=5	6
<=6	4
>6	1

Hrs/Wk Google Bins	Frequency
<=1	1
<=2	4
<=3	6
<=4	8
<=5	6
<=6	4
>6	1

LIS Degree	Hrs/Wk Google
Yes	3
No	5