User Skill Acquisition in Office Information Systems

Michael D. Cooper
School of Library and Information Studies, University of California at Berkeley, Berkeley, California 94720

The process of learning to use a new software package is incremental. Users begin learning one function and as the need arises, they acquire skills necessary to use another function. This article develops advent models of this process and tests the models on longitudinal data from more than 300 users of an office automation system called PROFS over a three-year period. The models explain the advent of use of certain PROFS commands in terms of the time since the study participant began using the system, and demographic characteristics such as job title, gender, years of work experience at the institution, and departmental affiliation within the organization. The models show that certain functions such as sending and receiving electronic mail are learned earliest, and that function use does depend on many of the demographic variables.

Introduction

This article proposes a model for investigating patterns of user interaction with information systems. Most information systems have a complex command or menu structure. While the user interface to the system may appear simple and intuitive, the primitive commands and options available to the user may be extraordinarily complex. One need only look at Personal Computer software packages such as Lotus 1-2-3 or WordPerfect to realize that a reasonably intuitive interface may lead the user through a considerable labyrinth to finally arrive at the command to be executed along with the parameters to be passed to the command.

One of the problems of evaluating interaction patterns is analyzing which commands or functions are used in a system out of the total available set. Given the available commands in the system, a related problem is to determine the chronology with which a user adopts functions and features of a software package. It is this later problem that is discussed in the article.

Several groups are interested in evaluating user interaction with an information system. These include the designers, trainers, managers, and users of systems. Designers need to know if the system is meeting the needs of the users and what kind of deficiencies there are to a particular interface and particular command structure. Trainers need to know what the user is actually doing with the system and, specifically, what kind of problems the user is having with using the system. Managers need information on the extent to which the system is meeting the needs of the users, whether usage warrants continuing to fund the system, and what the future plans for the system should be. Users need the functionality in the system that will let them perform their jobs in an efficient way.

Received October 26, 1990; revised February 13, 1991; accepted February 19, 1991.

Previous Studies

The problem addressed in this study is to analyze the skill acquisition sequence in using a software package. One approach, as Cambpell (1990) notes, is developmental: Human-computer interaction specialists need to perform longitudinal monitoring studies to understand the sequence with which individuals learn applications and how they become experts in the use of the system.

Information system usage can be studied by conjecturing about use patterns, by interviewing users, by making obtrusive or unobtrusive observations, or by capturing user interaction patterns through a logging mechanism and studying this log. This article takes the later approach, which requires good monitoring tools. Ferrari (1978), McKerrow (1988), and Svobodova (1976) both provide a review of monitoring tools and Schmidt and Pobuda (1988) summarize monitoring techniques and computer performance evaluation methods for a large bibliographic utility. Penniman and Dominick (1980) review the application of these tools to online information retrieval systems, and Larson (1986) and Tremain and Cooper (1983) describe monitoring tools for online catalogs and online information retrieval systems, respectively. Nielsen (1986) provides a more general view of the value of monitoring.

A number of studies have used monitor data to measure system and user performance but only one has

 CCC 0002-8231/91/010735-12$04.00

tried to develop a longitudinal analysis of user behavior. Crawford (1985), and Schmidt and Pobuda (1988) have reported system performance and usage characteristics for the RLIN bibliographic utility using monitor data. Cooper has analyzed the usage and response time of the MEDLARS system using monitoring data (1983a, 1983b).[1]

A majority of recent monitoring studies have focused on usage patterns of online catalogs. This is a particularly fruitful area of research since very little empirical evidence was available on how users performed searches until recently. Larson (1986) reports on a monitoring evaluation of the University of California's MELVYL system, Borgman (1985) reviews Ohio State University users' behavior patterns in searching their online catalog, Kaske (1988a, 1988b) analyzes searching behavior at the University of Alabama's online catalog,[2] and Larson (1990) analyzes subject searching behavior on the MELVYL system.

Tolle's (1983a, 1983b) work in analyzing online catalog use through transaction logs is the closest to the research discussed in this article. The portion of the project of interest here is that concerned with developing models of user interaction. Tolle's approach was to model the sequence of user interactions as a Markov process. Through transaction logs, Tolle was able to isolate discrete user sessions.[3] He developed a taxonomy which represented a mapping between online search system commands and metalevel *states* or supercategories of commands. Then he took transactions for a single search and mapped the commands into states. This process was repeated for a large number of searches.

A user begins in one state and moves to another. Tolle computed the probability of moving using the frequency data in the transaction logs and made estimates of the most likely paths that users will follow in issuing commands to perform an online search.

The present study takes a different approach to the same problem. It uses transaction log data to measure the first time (or advent) of a user starting to use the system. Then it calculates the probability of the user adopting the function into his or her repertoire of commands.

The Advent Model

An initial step in the analysis of the pattern employed by a user to learn a new system is to consider when a particular function or command is first used. For this purpose survivor analysis methods are employed. A *survivor function* is defined as determining the probability that an individual drawn at random from a set of users will continue to use the system for some amount of time t. In this article it is not the demise of an activity that is considered, but the advent of it. Hence, instead of calculating *survivor* functions,[4] this article calculates what might be called *advent* functions.

The procedure is to determine the proportion of users expected to have tried each function of a system by time t following the very first use of the system at time $t = 0$. These estimates are unconditional in the sense that they are not adjusted for the "competing hazard" of dropping out: some users i may not have tried function m by time t because they quit using the system altogether.

Advent functions can be estimated for any point in time, and the ratio between advent values at successive points in time (say a few months) allows one to discriminate between those activities picked up quickly by users (high ratio) to those picked up later (low ratio).

Probability of Function Use

Once it has been determined when the first use of a function has occurred, the next question is who uses which functions of a system in a typical work day? For each activity two use rates can be calculated. One is the *conditional* use rate, that is the share of all user days in which the activity appears for those users ever using the activity in a predefined period. The second is the *unconditional* use rate, the share of all user days in which the activity appears in any of the user-days in the sample period.

The incidence of use revealed by use rates can be analyzed as a random process in which the logarithm of the ratio of the probability of use of function i by user j on some use day t is a linear function of a vector of user characteristics X_j. These characteristics could be factors such as the job classification of the user, the administrative department to which the user belongs, and the user's age and/or gender. That is, if P_{ijt} is the probability that user j will undertake activity i in period t then,

$$\ln[(P_{ijt})(1 - P_{ijt})^{-1}] = X'_{jt}\beta_i, \qquad (1)$$

[1]While not reporting usage statistics, Lynch (1988) provides a number of comments on performance evaluation of online systems. These comments are a cautionary note reflecting a perspective in evaluating such measurement.

[2]In a more recent study, Turner, Kaske, and Baker (1990) used a primitive search logging mechanism and sophisticated heart rate monitoring system to measure, among other things, user anxiety in online searching.

[3]In itself, this is not a trivial task, since it is often difficult to determine when a user's session with an online catalog starts and when the next user's session begins. See Tremain and Cooper (1983) and Taylor, Taylor, and Edgerton (1989) for a discussion of this problem.

[4]See Kalbfleisch and Prentice (1980).

where β_i is a vector of coefficients.[5] The probability P_{ijt} is then

$$P_{ijt} = [1 + \exp(-X'_{jt}\beta_i)]^{-1} \quad (2)$$

Given a sample of use-days for which the incidence of activity i and user characteristics X_{jt} are recorded, the parameters of the expression above may be estimated using maximum likelihood techniques.

The Study Environment

The models were employed to evaluate the usage patterns of a software product called PROFS. PROFS is an acronym for *Professional Office System* and is sold by the IBM Corporation to users of their computer systems. PROFS provides an environment within which a user can, among other things, send and receive electronic messages, maintain a calendar or schedule, and prepare and file documents (word processing) (See: Using the Professional Office System (1988)).

The data about PROFS usage was collected in the context of a three-year study conducted at the University of California at Berkeley as part of a Joint Study with the IBM Corporation. The major purpose of the study was to analyze the feasibility of using an integrated office support system in an academic environment. The project involved IBM's installation of their 4300 series machines and numerous terminals and communication lines throughout five academic and administrative departments. These departments included the School of Library and Information Studies, the Department of Economics, the School of Business Administration, the Graduate School of Public Policy, and the Office of the Chancellor of the University. More than 300 individuals participated in the study by using the PROFS system.[6]

Descriptive Data about Function Use

The PROFS system appears to the user as a hierarchy of menus. In the Berkeley campus implementation of PROFS, there were three main menus of selectable activities (See: Using PROFS at Berkeley (1984)). Table 1 summarizes the usage of main menu choices. These data are for a sample of use conducted during the Spring of 1985. The data show that once having selected a certain function, users tend to perform it a number of times. For example, there are almost 3.5 calendaring actions for every one main menu invocation of this function. While the search action is selected only about five percent of the time from the main menu, about sixteen percent of all actions were searches. And because there were three main menus in the Berkeley Campus implementation of PROFS, one of the most frequently occurring main menu actions was simply transferring between main menus. This action, marked "transfer between main menus" in Table 1 constituted about twelve percent of all main menu actions.

PROFS operated under the CMS Operating System at Berkeley and one of the facilities of PROFS is to execute CMS commands such as FILELIST (list the names of files on one's account) or XEDIT (invoke the CMS editor to process a file). The most frequently occurring activity by users was to execute a CMS command. One explanation for this may be the nature of the study environment. Participants were given terminals and a CMS account with PROFS access for which they were not billed. Users may have found out that they could save money by simply doing their normal CMS computing for free on their PROFS account rather than at cost on their regular account, and the high "PROFS or CMS Command" usage may reflect that. A second explanation for the high execution of CMS commands within PROFS is that users may have found that the PROFS environment was good for some office automation activities and not as good for others. For example, they may have chosen to do their editing and document formatting outside PROFS in order to have what they considered more flexibility in document handling.

The main conclusion to be drawn from Table 1 is that although a wide variety of functions were available under PROFS at Berkeley, users selected only a very few for heavy use.

Detecting Tourism in System Use

Curious new users of any system tend to take "tours" of it to find out what it has to offer. Since the concern of this evaluation was to determine when the *advent* of use of a particular PROFS function occurred, it was necessary to specify what constituted "tourism" within the hierarchy of the PROFS menus and what constituted genuine initiation of work using a particular function. A *restricted* definition of what constituted *use* was developed mainly by requiring actions to take place once the user arrived at a particular menu. Examples of *use* of a task include sending an electronic message versus looking at the screen which lets you send a message, or preparing a document versus moving to the screen that allows a document to be prepared, but never performing any action on it.

Restricted definition counts were made for 11 major PROFS functions: (1) opening mail, (2) document search, (3) using calendar, (4) processing a mail log, (5) sending a note, (6) looking at a note log, (7) prepar-

[5]Instead of assuming that P_{ijt} is a function of X_{jt} and β_i, the logit transformation $\ln[(P_{ijt})(1 - P_{ijt})^{-1}]$ is used. The logit equation has the benefit of being able to better model changes in proportions than just P_{ijt} itself. See Snedecor (1989) for an elementary discussion of the benefits of logit transformations and Kalbfleisch and Prentice (1980) for details in this context.

[6]See Cooper (1990) for details on how participants were selected for participation in the study.

TABLE 1. Frequency distribution of IBM PROFS system function use, Spring 1985 sample,

Function Name	Main Menu Uses	Percent Distribution	Total Uses	Restricted Definition Uses	Percent Distribution
PROFS or CMS command at command line	35,550		35,550		
Review in basket	19,024	19.099	52,061	15,809	21.444
Transfer between main menus	12,117	12.165	12,117		
Search for document in database	4,907	4.926	18,746	11,596	15.729
Send a note to another user	4,507	4.525	9,635	3,810	5.168
Perform calendar functions	4,015	4.031	13,947	7,629	10.348
Manage mail log	3,526	3.540	9,713	5,087	6.900
Other actions	3,518	3.532	3,518		
Prepare document	3,284	3.297	3,978	9,658	13.100
Exit PROFS	3,211	3.224	3,211		
Manage note logs	1,657	1.664	4,247		
File documents	1,460	1.466	2,977	11,367	15.418
View personal direct. of names and numbers	668	0.671	1,558	367	0.498
View central directory of names and numbers	656	0.659	656		
Set a reminder	607	0.609	1,618	852	1.156
Send immediate message to another user	437	0.439	502	360	0.488
Manage distribution lists	138	0.139	138		
Create author profiles for use in document preparation	100	0.100	100		
Help	89	0.089	89		
Update names	68	0.068	68		
Change password	68	0.068	68		
Total	99,607	100.000	118,982	73,723	100.000

ing a document, (8) filing a document, (9) setting an automatic reminder, (10) reviewing one's personal directory, and (11) sending a message. Counts using the restricted definition are a subset of total uses.

Advent Function Analysis

Table 2 presents two points on the estimated advent functions for the first use of PROFS for each of eight activities and for dropping out. The first is at six months after the user first logged onto the system, and the second is at the first anniversary of initial logon. The ratio of the advent rate at six months to the advent rate at twelve is calculated.[7]

The results indicate that mail functions are used early. The advent ratio for "Review In-Basket" (i.e., look to see what kind of electronic mail you have) is .93. Functions like "Search for Document" (locate, in your account, a document you previously prepared) and "Manage Personal Directory" are learned late. The advent ratio for "File a Document" is .64. Some functions are not investigated for a long time, if ever: only one in five users is expected to perform a personal directory task after 12 months.

Conditional Use Rates

Given the advent of first use, what are the conditional and unconditional function use rates?[8] In Table 3 these rates are calculated for the 89 sample days during which there were 19,605 user-days of activity.[9]

Table 3 shows, for example, that there were 18,773 days of use observations for participants who opened mail at least once. There are also 832 days of use observations for users who never opened mail during the 89-day period. The unconditional use rate is 832 use-days divided by 19,605 total use days (18,773 plus 832 use days accumulated by those who never received a message). The conditional use rate is 3,427 use days observed with a mail opening divided by 18,773 (19,605 − 832).

[7]See "The Advent Model" Section.

TABLE 2. Advent rates, IBM PROFS activities, sample of 210 participants.[a]

PROFS Function	Advent Rate 6 Months After First Logon (1)	Advent Rate 12 Months After First Logon (2)	Ratio (1)/(2)
Review in-basket	.877 (.025)	.944 (.020)	.93
Perform calendar functions	.353 (.034)	.397 (.037)	.89
Send a note	.706 (.034)	.798 (.032)	.88
Manage note logs	.568 (.036)	.711 (.037)	.80
Set a reminder	.250 (.032)	.324 (.036)	.77
Prepare a document	.472 (.037)	.630 (.038)	.75
Manage mail log	.531 (.037)	.718 (.037)	.74
Search for document	.458 (.037)	.662 (.039)	.69
Send immediate message	.340 (.035)	.495 (.040)	.69
File documents	.180 (.028)	.280 (.035)	.64
End logons (quit use)	.145 (.025)	.231 (.030)	.63
Manage personal directory	.093 (.022)	.208 (.032)	.45

[a]Numbers in parentheses are standard errors of estimates. Estimation based on "Life Table" techniques; see Kalbfleisch and Prentice (1980) and Cooper (1990) for details.

TABLE 3. PROFS actions counts, 19,650 user-days.

Activity	Days Without Activity Occurring (1)	Days With Activity Occurring (2)	Total Use Days (1) + (2)	Nonuse Days (4)	Use Rate (Conditional) (2)/(3)	Use Rate (Unconditional) (1)/(1) + (4)
Review in-basket	15,346	3,427	18,773	832	0.18	0.17
Search for document	12,101	870	12,971	6,634	0.07	0.04
Calendar	7,089	709	7,798	11,807	0.09	0.04
Manage mail log	13,527	798	14,325	5,280	0.06	0.04
Send a note	14,247	1,766	16,013	3,592	0.11	0.09
Manage note logs	13,834	728	14,562	5,043	0.05	0.04
Prepare a document	12,002	1,080	13,082	6,523	0.08	0.06
File documents	5,477	442	5,919	13,686	0.07	0.02
Set a reminder	6,819	304	7,123	12,482	0.04	0.02
Personal directory	5,237	81	5,318	14,287	0.02	0.00
Send immediate message	9,676	213	9,889	9,716	0.00	0.01

[8]See the definition of these concepts in earlier section "Probability of Function Use."
[9]The sample is limited to participants who were active for at least one of the days in the sample period.

TABLE 4. Definitions of variables used in Table 5.

Variable	Definition
INTERCPT	Intercept: Always = 1.
OS_AGE	Time in years between beginning of use of PROFS and first use of the function
OS_AGE2	OS_AGE squared
Job Type	
EXECTIVE	Executive = 1 if participant is Executive, = 0 otherwise
OFFICER	Program Officer = 1 if participant is Program Officer, = 0 otherwise
ADM_ANL	Administrative Analyst = 1 if participant is Admin. Analyst, = 0 otherwise
GEN_ANL	General Analyst = 1 if participant is General Analyst, = 0 otherwise
FAC_SEC	Faculty Secretary/Administrative Assistant = 1 if participant is Faculty Secretary, = 0 otherwise
STU_SEC	Student Affairs Secretary = 1 if participant is Student Secretary, = 0 otherwise
FACULTY	Faculty = 1 if participant is Faculty, = 0 otherwise
OTHERJOB	Other = 1 if participant is Other, = 0 otherwise
ADM_SEC	Administrative Secretary = 1 if Admin. Secretary, = 0 otherwise
ADM_FAC	Administrative Faculty = 1 if participant is Admin. Faculty, = 0 otherwise
Unit	
ECO	Department of Economics = 1 if participant is in Economics, = 0 otherwise
LIS	School of Library and Information Studies = 1 if participant is in Library and Information Studies., = 0 otherwise
SBA	School of Business Administration = 1 if participant is in Business Admin., = 0 otherwise
SPP	Graduate School of Public Policy = 1 if participant is in Public Policy, = 0 otherwise
Personal Characteristics	
MALE	= 1 if participant is male, = 0 otherwise
AGE	= age in years
AGE2	= age squared
UCBTIME	= years as Univ. of California, Berkeley employee
UCBTIME2	= UCBTIME squared
STARTLOG	= time in years between beginning of project and first use of PROFS
DWRITER	Displaywriter = 1 if participant used an IBM Displaywriter for access to PROFS, = 0 otherwise

Logit Estimates of Probability of Use

Equation (2) showed that the probability (P_{ijt}) of use of a system feature can be estimated by a function of user characteristics, X_j. The characteristics chosen as a basis for calculating the probabilities are given in Table 4.[10] They include the job title of the individual using the system, the departmental affiliation of the individual, the person's gender and age, as well as the number of years the person has been employed at the University.

Logit equations are then used to estimate the probabilities in equation (2) for each of eight PROFS functions.

[10]These are similar characteristics to those used in the Hazard analysis presented in Cooper (1990).

First, the "goodness of fit" of the functions in Table 5 is considered, then the results are reviewed on the basis of the independent variables.

Goodness of Fit. This Subsection first concentrates on the statistical significance of the results; the next Subsection gives the numerical implications. Since logit equation (2) produces probability estimates, but what is observed are discrete events, it is impossible for the equation to "fit" perfectly in the sense that every use of the system will be associated with an estimated probability of 1 and every user day in which the associated function is not used will have an estimated probability of precisely 0. Two measures are included in the analysis that provide a sense of how well the equation does. One is the "fraction of concordant pairs of predicted probabilities and responses," CONCPRS. CONCPRS is calculated by first ranking all observations by the esti-

TABLE 5. Logit estimation results, probability of function use.[a,b]

		Function							
		Review In-Basket		Search for Document		Calendar		Manage Mail Log	
Variable	Mean	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability
INTERCPT	1.00	−5.983	0.000	6.104	0.000	−20.447	0.000	4.092	0.000
OS_AGE	1.36	−0.248	0.122	−2.944	0.000	1.859	0.001	−1.734	0.000
OS_AGE2	2.17	0.249	0.000	1.046	0.000	−0.601	0.000	0.629	0.000
EXECTIVE	0.06	−1.103	0.000	−1.302	0.013	−0.585	0.101	−2.228	0.000
OFFICER	0.08	−0.041	0.738	−1.360	0.000	−1.056	0.000	−0.517	0.014
ADM_ANL	0.08	0.156	0.212	−0.299	0.178	−2.398	0.001	0.278	0.185
GEN_ANL	0.07	0.273	0.022	0.356	0.050	−0.101	0.746	−0.153	0.429
FAC_SEC	0.12	0.804	0.000	−1.763	0.000	−0.982	0.001	0.221	0.238
STU_SEC	0.06	0.212	0.085	−0.408	0.114	−1.387	0.000	−0.412	0.103
FACULTY	0.02	1.136	0.000	0.653	0.542	−1.041	0.006	−2.929	0.005
OTHERJOB	0.07	−0.976	0.000	−1.934	0.000	−0.325	0.416	−0.359	0.214
ADM_SEC	0.13	0.043	0.677	−1.263	0.000	2.592	0.000	−0.514	0.006
ADM_FAC	0.23	−0.043	0.731	−0.055	0.824	−1.708	0.000	−1.073	0.000
ECO	0.22	0.838	0.000	−1.788	0.000	0.508	0.043	−1.164	0.000
LIS	0.13	1.921	0.000	−1.579	0.000	0.052	0.830	−1.522	0.000
SBA	0.22	−0.590	0.000	−1.879	0.000	1.566	0.000	−0.984	0.000
SPP	0.09	−0.397	0.000	−0.233	0.150	0.685	0.007	−1.172	0.000
MALE	0.42	−0.447	0.000	−0.680	0.000	0.354	0.121	−0.155	0.257
AGE	41.82	0.224	0.000	−0.141	0.005	0.799	0.000	−0.167	0.000
AGE2	1,851.71	−0.002	0.000	0.001	0.124	−0.009	0.000	0.002	0.000
UCBTIME	10.38	0.015	0.206	−0.222	0.000	0.076	0.020	−0.076	0.000
UCBTIME2	170.71	−0.002	0.000	0.005	0.000	−0.005	0.000	0.001	0.099
DWRITER	0.22	0.064	0.384	−0.472	0.001	−2.939	0.000	−1.282	0.000
CONCPRS		0.753		0.827		0.828		0.764	
RNKCORR		0.521		0.679		0.674		0.564	
Number of observations		12,242		8,535		5,194		9,440	

Table 5. (continued) Logit estimation results, probability of function use.[a,b]

		Manage Note Logs		Prepare a Document		Send Immediate Message		Send a Note	
Variable	Mean	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability	Coefficient	Probability
INTERCPT	1.00	−10.133	0.000	−3.946	0.000	−3.900	0.086	−9.273	0.000
OS_AGE	1.36	−2.066	0.000	0.648	0.021	−1.899	0.002	−1.158	0.000
OS_AGE2	2.17	0.657	0.000	−0.146	0.091	0.640	0.000	0.547	0.000
EXECTIVE	0.06	−0.857	0.022	1.174	0.000	−1.977	0.098	−0.438	0.089
OFFICER	0.08	−0.131	0.501	0.460	0.010	−0.596	0.352	−0.090	0.602
ADM_ANL	0.08	0.301	0.171	1.304	0.000	−0.944	0.173	0.004	0.981
GEN_ANL	0.07	−0.297	0.169	0.562	0.006	−0.313	0.455	0.624	0.000
FAC_SEC	0.12	0.996	0.000	0.882	0.000	1.486	0.000	2.028	0.000
STU_SEC	0.06	−0.437	0.100	0.380	0.051	−0.761	0.275	0.102	0.000
FACULTY	0.02	−0.418	0.301	−1.467	0.018	1.489	0.023	1.422	0.000
OTHERJOB	0.07	−0.332	0.282	−1.107	0.000	0.933	0.103	0.292	0.106
ADM_SEC	0.13	0.787	0.000	−0.119	0.528	0.594	0.104	1.132	0.000
ADM_FAC	0.23	−0.012	0.960	−0.576	0.003	1.137	0.031	0.838	0.000
ECO	0.22	0.105	0.429	0.359	0.004	−0.532	0.087	0.825	0.000
LIS	0.13	0.141	0.335	−0.155	0.250	−0.760	0.034	1.361	0.000
SBA	0.22	−0.981	0.000	−0.902	0.000	−1.440	0.000	−1.009	0.000
SPP	0.09	−0.314	0.104	−0.501	0.001	−1.587	0.000	−0.336	0.015
MALE	0.42	−0.889	0.000	0.570	0.000	−0.590	0.136	−0.683	0.000
AGE	41.82	0.396	0.000	0.114	0.001	−0.007	0.952	0.308	0.000
AGE2	1,851.71	−0.004	0.000	−0.001	0.000	0.001	0.646	−0.003	0.000
UCBTIME	10.38	0.109	0.000	−0.195	0.000	0.408	0.000	0.102	0.000
UCBTIME2	170.71	−0.006	0.000	0.005	0.000	−0.024	0.000	−0.005	0.000
DWRITER	0.22	−0.500	0.000	−0.772	0.000	−0.178	0.510	−0.512	0.000
CONCPRS		0.704		0.734		0.733		0.754	
RNKCORR		0.451		0.495		0.542		0.527	
Number of observations		9,727		8,803		6,602		10,499	

[a]"Probability" column gives probability of estimation of corresponding coefficient with same or greater absolute size given observation count when the true coefficient value is zero. CONCPRS and RNKCORR are defined in the text. The dependent variables are defined as follows: Review In-Basket: User looks at note or document in in-basket; Search for Document: User performs search for document; Calendar: User looks at calendar; Manage Mail Log: User performs some operation with maillog; Manage Note Logs: User reviews a note log; Prepare a Document: User enters editing facility and executes some task related to document processing; Send Immediate Message: User sends "immediate" message; Send a Note: User sends a note. See Table 4 for definitions of the independent variables.

mated probability of use of the indicated function. Then the probability is compared with the outcome. All cases in which the function was used and the estimated probability exceeded .5, or in which the function was not used and the estimated probability was less than .5, are designated "concordant." The CONCPRS value of .753 for the "Review In-Basket" function means that 75% of all cases were correctly predicted on this criterion. Among the functions for which logits were estimated, the best prediction was achieved for calendar use, the worst for "Manage Note Logs."

An alternative measure of goodness of fit is RNKCORR, the rank correlation between estimated probability of function choice and actual outcome.

Probability Estimates. Turning to the independent variables themselves, the large number of observations allows considerable precision in estimation. Here the estimated probability of observing the association between the outcome and the independent variable that appears in the sample is reported, rather than standard errors for the estimated coefficients. This estimated probability is conditional in that, with repeated sampling, the observed correlation will disappear; that is, under the null hypothesis the true value of β_i is zero. The analysis focuses on the relationships for which these probabilities are less than an arbitrary value of 0.10.

The variable "Office System Age" (OS_AGE) in the equation represents the time in years between the beginning of use of the PROFS system and the first use of the function. For all the models reported in Table 5, Office System Age is entered with a quadratic term to allow for nonlinearity of change in use with the passage of time. For all eight functions the OS_AGE effects are jointly significant. The remaining functions fall into three categories. For everything but calendar use and document preparation, use initially declines with time. However, the positive coefficients on the square of OS_AGE indicate that this decline slows and, in some cases, reverses direction of change within the domain of OS_AGE included in the data. For the "Review In-Basket" function, this reversal occurs after almost exactly one year of experience with the system. For the "Manage Note Logs" function, the estimated reversal point is at three years, almost outside the domain of the data we study. Since the quadratic specification must be only an approximation to the "true" functional relationship between OS_AGE and note log management, the results are probably telling us only that the incidence of this activity declines at a decreasing rate with time on the system. The third pattern is represented by "Prepare a Document;" here activity increases with office system experience, but the rate of increase declines with time.

In this article, the use of a PROFS function is measured relative to what is expected from a general secretary in the Chancellor's Office. Members of most of the occupational categories perform the "Review In-Basket" function more frequently than members of this reference group. The only job classifications in which members do significantly less mail handling are executives and people in the "other" group. Likewise, members of all other classifications save one either look for documents with less or about the same frequency as the reference group; most appear to open it less frequently. The exception is, again, the general analyst. Interestingly, administrative secretaries appear much more likely to use the calendar function than is true for the general secretary in the Chancellor's Office, and the administrative analysts in the sample appear to have used the calendar function extremely infrequently. Other results are generally consistent with expectations except that, aside from faculty, administrative secretaries, and research assistants, everyone seems to be more likely to prepare documents using PROFS than a general secretary. At the time the study was conducted, this was an interesting result since personal-computer-based word processing was not as prevalent as it is now and there were less choices available to the general secretary for word processing. Nevertheless, PROFS was not the system of choice for this category of user for this function.

The differences in use patterns between the participating units are again significant and varied. For activities related to accessing the PROFS database (Search for a Document; Manage Mail Log), function use by users in other units is consistently less than for the Chancellor's Office. For other functions, use in the Economics Department and in the School of Library and Information Studies tends to be greater, although in some cases (e.g., Manage Note Logs) the differences are not statistically significant. The compact quarters that house the staff of the School of Library and Information Studies apparently foster communication of all sorts, including the transmission of lots of PROFS notes.

Differences in function use by gender do appear. All other things being equal, men are less likely to undertake any operation involving note or message sending than are women; they are less likely to conduct database-related activities; and they are more likely to use the document preparation functions.

Other things being equal, older participants were more likely than young ones to send and receive mail, use the notelog, use the calendar, and prepare documents. They were less likely to undertake activities involving the PROFS database. Age being constant, individuals who have worked on the Berkeley Campus for a long time (as measured by UCBTIME) were less likely to use all functions in the system except those associated with note and message sending and receiving functions.

About one-fifth of the potential use days in the sample were accumulated for operators using the IBM Displaywriter (a stand-alone word processor) as a terminal. The last row in Table 5 shows that Displaywriter

operators were less likely than others to do all PROFS activities except open mail. Clearly Displaywriter functions were substituted for PROFS. And it appears that use of the Displaywriter made it less likely the user would employ other PROFS functions, such as calendaring, even when they were not available on the workstation.

Probability of Activity Occurrence

Sample probabilities of the occurrence of an activity calculated from the data in Table 5 are presented in Table 6. Calculations begin by assuming the baseline General Secretary in the Chancellor's Office who is 35 years old, has one year of office systems experience, and has worked for the University for 10 years. Consider the effect on the likelihood of opening mail if these characteristics are changed. The baseline probability is just .21. While experience at the University may have a statistically significant effect on the likelihood of opening mail, the consequences are not numerically very large: 50% less experience raises the likelihood of receiving mail by about .01. Mail use in the Graduate School of Public Policy is less frequent by a third, but, as we have already seen, secretaries in the School of Library and Information Studies are more than three times more likely to receive mail. Even a faculty member in the School of Library and Information Studies receives mail more frequently than is true for the general secretary in the Chancellor's Office.

While secretaries in the School of Library and Information Studies are much more likely to send and receive notes than is true for secretaries in the Chancellor's Office, calendar use appears much greater among secretaries in the School of Public Policy. Nonetheless, most of these numbers are small; a probability of calendar use of .138 means on average a general secretary in the Graduate School of Public Policy looks at her calendar only every seven working days. What is more likely, of course, is that within each category the average masks wide dispersion: some use it frequently, some use it not at all. Similarly low use rates are apparent for the document preparation function.

Hence while the multivariate analysis detects interesting variations among users in the degree of participation in the study, the conclusion remains the same as was drawn from the first tables: only a few functions, such as reviewing one's in-basket, were used heavily, and the rest were used only occasionally.

Summary and Conclusions

Users of information systems do not take advantage of all the functions and features of an information system the moment they are introduced to it. Instead, users begin with one or two functions and gradually learn about additional ones. This article has presented a methodology for modeling this process using advent analysis. The models formulate the probability of selecting a particular command as a function of the time since first exposure to the system, as well as organizational and demographic characteristics of the users.

The models were tested with empirical data from a longitudinal study of over 300 users of the IBM PROFS Professional Office System on the University of California at Berkeley campus for a three-year period. Descriptive data from that study showed that a few functions of the system were used very frequently

TABLE 6. Probability of activity occurrence.

	Review In-Basket	Search for Document	Calendar	Manage Mail Log	Send a Note	Manage Note Logs	Prepare Document	Send Immediate Message
General Secretary, Chancellor's Office	0.210	0.208	0.118	0.192	0.077	0.103	0.069	0.052
General Secretary, Chancellor's Office, five years at UC Berkeley	0.221	0.357	0.116	0.243	0.069	0.093	0.117	0.040
General Secretary, School of Public Policy	0.152	0.172	0.209	0.069	0.057	0.078	0.043	0.011
General Secretary, School of Library and Info. Studies	0.645	0.051	0.123	0.49	0.246	0.117	0.060	0.025
Faculty Secretary, School of Library and Info. Studies	0.802	0.009	0.050	0.061	0.713	0.264	0.134	0.101
Faculty Member, School of Library and Info. Studies	0.850	0.094	0.047	0.003	0.575	0.080	0.015	0.101

(such as managing electronic mail and maintaining schedules and calendars) and most were used infrequently (Table 1). Advent rates were calculated for the functions available in the system. These rates were computed six, and then 12 months after the first logon to the system by the users in the study. It was found that the functions of the system that are used earliest are the same as those used most frequently (managing electronic mail and maintaining calendars), and there is a definite hierarchy in the sequence with which functions are used (Table 2). This hierarchy is important for the designers, trainers, managers, and users enumerated earlier. Designers can see that, for example, the "Send Immediate Message" function that might have appeared to be highly relevant, was not adopted early in the user's career with the system. Likewise, functions that might have appeared to be highly relevant were just that—the calendar function.

Trainers can concentrate their first courses in how to use the system on the functions they know will be the most used, and defer other teaching tasks until later, provided they have information like Table 2 available. Tutorial materials can be structured in the same way.

If managers are concerned about the usefulness of a system this kind of information can be used to tell them the answers. Managers who believe that their employee productivity can see that it takes a long time, for example, before the PROFS system is employed to electronically file documents. Managers may want to ask whether the system is suited for the task or whether the employees have not received adequate training in the use of this function.

Logit estimates were made to determine how the probability of using a particular function varied with organizational and demographic characteristics of the user as well as the length of time between first logon to the system and use of the function. The last variable was consistently significant across all functions. The probability of the use of a function at a point in time was found to vary by job classification and by organizational unit. For certain functions, the age of the individual and the person's gender made a difference in explaining when that person was likely to start using the command.

The article has shown that there is a healthy payoff in an investment in a logging system and advent models. The logging system need only record the date and time that a user performs a specific command or function. With that information, advent models can be employed to provide valuable evaluative information about system use patterns.

Acknowledgments

I would like to thank Michael Wiseman for his help in the development of the models and to Karl Iorio for his excellent programming assistance and general advise throughout this project. This research was supported in part by a Joint Study Agreement between the IBM Corporation and the University of California at Berkeley.

References

Borgman, C. L. (1985). End user behavior on an online information retrieval system: A computer monitoring study. In J. J. Kuehn (Ed.), *International Conference on Research and Development in Information Retrieval* (pp. 162–176). 6th Annual International ACM SIGIR Conference. Association for Computing Machinery, New York, pp. 162–176.

Campbell, R. L. (1990). Developmental scenario analysis of smalltalk programming. In J. Chew & J. Whiteside (Eds.), *CHI '90 Conference Proceedings* (pp. 269–276). Seattle, Washington April 1–5, 1990. Association for Computing Machinery, New York.

Crawford, W. (1985). The RLIN command analysis system: Measuring use and performance of an online system. *Information Technology and Libraries, 4,* 29–51.

Cooper, M. D. (1990). Longevity and intensity of office system use. *Journal of the American Society for Information Science, 42,* 644–656.

Cooper, M. D. (1983a). Response time variations in an online search system. *Journal of the American Society for Information Science, 34,* 374–380.

Cooper, M. D. (1983b). Usage patterns of an online search system. *Journal of the American Society for Information Science, 34,* 343–349.

Ferrari, D. (1978). *Computer systems performance evaluation.* Englewood Cliffs, NJ: Prentice Hall.

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data.* New York: John Wiley and Sons.

Kaske, N. K. (1988a). A comparative study of subject searching in an OPAC among branch libraries of a university library system. *Information Technology and Libraries, 7,* 359–372.

Kaske, N. K. (1988b). The variability and intensity over time of subject searching in an online public access catalog. *Information Technology and Libraries, 7,* 273–287.

Larson, R. R. (1990). The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science, 42,* 197–215.

Larson, R. R. (1986). *Workload characteristics and computer system utilization in online library catalogs.* Doctoral Dissertation, University of California, Berkeley (University Microfilms No. 8624828).

Lynch, C. A. (1988). Response time measurement and performance analysis in public access information retrieval systems. *Information Technology and Libraries, 7,* 177–183.

McKerrow, P. (1988). *Performance measurement of computer systems.* Reading, MA: Addison-Wesley Publishing Company.

Nielsen, B. (1986). What they say they do and what they do: Assessing online catalog use instruction through transaction monitoring. *Information Technology and Libraries, 5,* 28–34.

Penniman, W., & Dominick, W. (1980). Monitoring and evaluation of online information system usage. *Information Processing and Management, 16,* 14–35.

Schmidt, J. G., & Pobuda, M. (1988). Capacity modeling at RLG. *Information Technology and Libraries, 7,* 173–177.

Snedecor, G.W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.) Ames, IA: Iowa State University Press.

Svobodova, L. (1976). *Computer performance measurement and evaluation methods: Analysis and application.* Computer Design and Architecture Series No. 2. New York: American Elsevier Publishing Company.

Taylor, C. M., Taylor, R. G., & Edgerton, J. G. (1989). Refinement of a method for determining the optimal interactive timeout in-

terval for OPAC terminals. *Information Technology and Libraries, 8,* 295–297.

Tolle, J. E. (1983a). *Final report to the council on library resources volume I. Current utilization of online catalogs: Transaction log analysis.* Dublin, Ohio: OCLC Online Computer Library Center, Inc., March 31, 1983. 127pp. (ED231402).

Tolle, J. E. (1983b). Transaction log analysis online catalogs. In J. J. Kuehn (Ed.), *International Conference on Research and Development in Information Retrieval* (pp. 147–160). 6th Annual International ACM SIGIR Conference. New York: Association for Computing Machinery.

Tremain, R., & Cooper, M. D. (1983). A parser for on-line search system evaluation. *Information Processing and Management, 19,* 65–75.

Turner, P. M., Kaske, N. K., & Baker, G. S. (1990). The effect of baud rate, performance anxiety, and experience on online bibliographic searches. *Information Technology and Libraries, 9,* 34–42.

Using the professional office system (1988). Publication Number SH20-5604. International Business Machines Corporation.

Using PROFS at Berkeley (1984). Document Number IBM 1.3.5 Release 2.1. Berkeley, CA: University of California, Berkeley: Office of Computing Affairs.