1989

USE AND PRODUCTIVITY IN PERSONAL COMPUTING: AN EMPIRICAL TEST

Brian T. Pentland
Massachusetts Institute of Technology

USE AND PRODUCTIVITY IN PERSONAL COMPUTING: AN EMPIRICAL TEST

Brian T. Pentland
Sloan School of Management
Massachusetts Institute of Technology

ABSTRACT

This paper provides some empirical evidence on the link between computer use and the efficiency and effectiveness of voluntary, direct users of personal computers. A survey of accounting professionals in the Internal Revenue Service (N = 1110) provides self-reported levels of use, efficiency and effectiveness, as well as data on training, management policy and user characteristics. A second survey of completed audits (N = 1851) provides an objective standard of comparison for the self-reported use and performance data.

Although users believe the computer is faster for some tasks, it does not improve their overall efficiency. The divergence is accounted for in part by managerial policies that encourage use of the system for marginal tasks and in part by users choosing to use the tool for activities that could be done more quickly manually. Data on the association between use and effectiveness show that while users believe the computer makes them more effective, much of this perceived value is symbolic rather than substantive. Users benefit from a sense of professionalism and self-esteem, but it is not clear whether the organization as a whole benefits.

The low association between use and productivity suggests that researchers should resist the temptation to regard use as a proxy for implementation success in the absence of actual productivity measures. Practitioners should be aware that policies which promote use may actually hurt productivity by encouraging users to apply technology to tasks where it is only marginally useful.

1. INTRODUCTION

Managers and other professionals are creating more applications and using more computers in their work (Benson 1983; Guimaraes 1986), but are they becoming more productive as a result? This important question has no clear answer. Senior management has grown wary of indiscriminant use of IS technology (Sutherland 1988), and econometric data suggest that, in recent years, investment in computing resources has not contributed to overall economic performance (Loveman 1987). Although IS researchers emphasize that the link between use and productivity should not be taken for granted (Trice and Treacy 1988), there has been very little empirical research on the productivity effects of end user or personal computing.

This paper addresses this gap in the literature by providing data on use, efficiency and effectiveness in the context of a personal computing system used by accounting professionals in the Examination Division of the United States Internal Revenue Service (IRS). The IRS is a useful context to study this issue because

1. over 14,000 professionals are employing personal computers in their work on a discretionary basis;

2. the task and the technology are controlled for by the organization;

3. there are clear, measurable standards of performance in the work; and

4. both subjective and objective data can be collected.

Given this unique research context, it is possible to test the relationship between use and productivity for professionals who are voluntary, direct users of personal computers.[1]

2. EXISTING LITERATURE ON PERSONAL COMPUTING AND USER PRODUCTIVITY

Because of the pragmatic difficulty of obtaining valid measures, very few studies attempt to measure the productivity impacts of information systems in field situations. Fudge and Lodish (1977) performed a quasi-experiment in a sales organization using ten matched pairs of sales representatives. These subjects were "direct" users (Lefkovits 1979) of a call planning decision support system. After six months, system users averaged 8.1 percent higher bookings than their non-computerized counterparts. Lucas (1975) also analyzed objective performance measures for a sales force using computerized reports to plan their sales activities. The subjects of this study were "indirect" users (Lefkovits 1979), since they received prepared reports over which they had no personal influence. Robey's (1979) study of MIS use by an industrial sales force revealed a high correlation between two measures of system use and self-reported job performance and performance visibility. In Crawford's (1982) case study of an electronic mail system, users reported increased personal productivity and managerial effectiveness. While this study clearly links use and productivity, it does so for a specialized kind of application which is qualitatively different from typical end user computing, and relies on self-reported measures of performance. Only Benson (1983) reports any data specifically on productivity improvements from microcomputers, but these findings are limited to three anecdotes. While these findings support our intuition that end users of computing technology *can* be more productive, they barely begin to elaborate on the conditions under which end users *will* be more productive.

The relative lack of empirical research may be attributable to the belief that the distinction between use and productivity is unnecessary for voluntary end users (Cheney, Mann and Amoroso 1986; Robey 1979; Maish 1979). The reasoning is that, unless use is mandatory, users will choose to apply computing technology only to those tasks where it is clearly productive. If this were true, then voluntary PC users would be productive in all that they do with their PCs. At a minimum, two assumptions are needed to make this argument work. First, users must have goals congruent with those of the organization so that their choices regarding computer use always reflect the interests of the organization. In terms of agency theory (Jensen and Meckling 1976), the organization would have no agency costs with respect to computing resource decisions. Second, users must be rational actors, able to correctly assess the productivity implications of their computing resource decisions. Given these assumptions, use becomes a reasonable proxy for productivity (or whatever the organizational goal might be).

While rational economic behavior is certainly part of organizational life, it is by no means all of organizational life (c.f., Burrell and Morgan 1979). Like any other aspect of organizational behavior, decisions by end users are subject to political and symbolic influences which are not constrained by economic rationality. In the context of systems acquisition, design and implementation, the importance of organizational politics has been well documented (Pettigrew 1973; Markus 1983; Robey and Markus 1984). Political behavior results from conflict of goals or interests within the organization, which implies a divergence from the assumption of goal congruence. Information systems also provide important symbolic value to users, who may elect to computerize some activity because doing so confers a certain aura of prestige or professionalism (Feldman and March 1981; Myer 1982; Kling and Gerson 1977). Decisions made for symbolic reasons may not be economically efficient, in violation of the second assumption.

Quite apart from these "irrational" behaviors, end-users may still violate the assumptions necessary to make use a reasonable proxy for productivity. For example, users may not be able to evaluate the efficacy of a particular computing application (this may be especially true if the application involves significant programming). Given these considerations, the assumption that use per se is indicative of actual productivity improvements is difficult to justify.

This realization, coupled with the difficulty of measuring actual productivity changes, has led some researchers to develop alternative measures of system effectiveness. "User information satisfaction" (Ives, Olsen and Baroudi 1983; Bailey and Pearson 1983) has been proposed as a measure that is particularly appropriate when use is mandatory. By asking users how well a system meets their information requirements for decision-making (e.g., timeliness, relevance, ease of use), these instruments indicate the respondents' subjective impressions of system effectiveness. In many cases, these user impressions will be the best possible indicators of actual system effectiveness.

The value of user information satisfaction as an indicator of system effectiveness may be limited to cases where a system is used primarily to make decisions, rather than doing some other kind of work. Analytical and decision support tasks represent only a fraction of the applications to which end user and personal computing is applied (Rockart and Flannery 1983; Guimaraes 1986). This observation clearly applies to the end users in this study, whose work includes correspondence, legal research, report preparation and other production-oriented tasks (described in more detail below). In this kind of situation, "information satisfaction" is irrelevant because the system is not providing the user with any information. More diffuse measures of user satisfaction (which do not isolate specific dimensions such as relevance and timeliness) may capture the symbolic and political influences discussed above rather than anything related to effectiveness or productivity (e.g., "I really like this system because it cost a lot of money and it really enhances my prestige"). In general, user satisfaction may be the best indicator of system effectiveness available in many situations, but it should be used carefully and not assumed to indicate actual productivity improvements.

3. A CONTINGENCY MODEL OF USER PRODUCTIVITY

The position taken here is that the link between computer use and user productivity is contingent on a variety of factors and worthy of empirical examination (Trice and Treacy 1988). The model presented here builds on previous models by focusing on the discretion that voluntary users have over where and how to apply computer technology. By emphasizing user discretion and productivity (rather than system success), this model attempts to deal explicitly with the fact that personal computers (and the enormous library of software available for them) are not a single technology and are not applied to a single task. Rather, personal computers represent a range of tools and can be applied to a range of tasks. To correctly conceptualize the effect of the personal computer on productivity, the whole range of tools and tasks must be considered.

Lucas (1975) presented a model of use and performance (defined as efficiency and effectiveness) which identified personal and situational factors as important. Lucas hypothesized that performance can be either positively or negatively related to IS use, depending on the relevance of the information provided by the system. While this model was restricted to a specific task-technology pair, this hypothesis highlights the critical importance of the fit between task and technology in the creation of useful, productive tools (Goodhue 1986; DeSanctis and Gallupe 1987). Not all applications of computer technology are equally productive. For example, an electronic spreadsheet can be used for all kinds of computations, large and small. But while a spreadsheet is enormously helpful for a large computation that needs to be frequently updated with new data, it is only marginally helpful for adding a small column of numbers on a one-time basis. The concept of task-technology fit will be the starting point for the model.

Adequate user skills and motivation are required in order to realize the potential of a given technology in a given task. In principle, a word processor is ideally suited to writing reports and memos, but its actual efficacy is influenced by the ability and motivation of the user to type. A similar observation can be made about any kind of tool, computer-based or not: user skills and motivation are necessary ingredients for productivity.

The relationship between skill and performance is simple and direct. Assuming motivation is held constant, performance can be used to measure skill (e.g., Peters 1977) and, for many purposes, performance practically defines skill (e.g., sports or music). However, the relationship between motivation and performance is more complex and can be entered into the model in two ways: as a predictor of use or as a predictor of effort. DeSanctis (1983) presents a model of motivation as a predictor of use and shows that the expectancy of desirable outcomes creates a motivational state that leads to use. Davis' (1985) model of technology acceptance also focuses on the link between user motivation and software use, as do a host of other studies. A separate stream of literature, also derived from expectancy theory (Vroom 1964), relates motivation to effort, which can lead to improved performance (Peters 1977; Campbell and Pritchard 1976).

In the model presented here, motivation will be subsumed under the use variable. This formulation follows the conceptual framework established by DeSanctis (1983) and Davis (1985) linking motivation to use and makes the simplifying assumption that if users are motivated enough to try a new technology, they are motivated enough to put forth a reasonable effort in using it. Although motivation is clearly antecedent to use and performance, it does not appear directly in the model.

The model addresses the fact that managerial and professional workers perform a wide range of tasks on a day to day basis by summing productivity changes over tasks. Although some tasks may be fairly routinized, professionals tend to lead a varied life. Faced with a variety of tasks to perform, voluntary users are constantly in the position of choosing when and how to apply computing technology. Furthermore, users have a choice about which specific technology to apply to a given task. For example, it may be possible to accomplish a given task with a spreadsheet, a database, or perhaps with pencil and paper. For the reasons discussed earlier, we cannot assume that users necessarily act to maximize their productivity in making these choices.

Taken together, these considerations lead to the simple conceptual model of user productivity shown in equation 1. Productivity in any given task depends on whether the computer is used, the fit between the task and the technology, and the skill level of the user. Overall productivity depends on the summation over the set of tasks performed.

$$\text{Productivity} = \tilde{o} \sum_{\text{tasks}} f\,(\text{Use}_i, \text{Fit}_i, \text{Skill}_i) \qquad (1)$$

This model is intended to emphasize the discretionary aspect of use. For whatever reasons, users choose where to apply their tools. One implication of this model is that if users make poor choices about where to apply technology, they can lower their productivity. Another implication is that even for rather poorly designed systems, there may still be a small range of tasks where they can still be put to productive use. Whether or not such uses are cost justified is another question, of course. The point here is that given a particular set of tools, a skill base and mix of tasks, the user's choices about where to apply the tools will determine the overall productivity outcome.

In the discussion that follows, this model will provide an interpretive framework for analyzing empirical results, but it will not be explicitly tested against an alternative model. The data collected allow a detailed examination of the relationship between use and productivity for a given set of tools, tasks, and user skill base. The effects of skill will be either controlled for implicitly or treated via proxy measures such as user experience and training. The question of fit is necessarily left implicit, since the set of tools and tasks are fixed by the organizational setting. Estimating these relationships would require detailed data on the performance implications of using alternative systems for each audit task. While the data presented here distinguish between manual and computer-based audits, and between the use of alternative system features for those audits, there is no distinction made between alternative features for specific tasks.

4. THE RESEARCH CONTEXT AND METHODOLOGY

The subjects of this study are accounting professionals in the Internal Revenue Service, called revenue agents, who conduct field audits of corporations, partnerships, sole proprietorships, and complex individual returns. Revenue agents are not managerial employees; they are "street level" personnel (Danziger and Kramer 1986) who actually conduct audits. Although tax auditing requires frequent decisions about the scope and level of detail to include in a given audit, it would be misleading to characterize revenue agents as decision makers. Their work involves examining books and records, legal research, preparation of reports, and computation of taxes. The work product of revenue agents consists entirely of audits of federal tax returns.

The system being used by the revenue agents in this study is called the Automated Examination System (AES). At the time of the study, AES consisted of a laptop portable computer and a variety of software for use in field audits. With the exception of access to commercial legal research services (which was contingent on the budget and policies of each district), the first phase of AES was completely stand-alone: one agent, one computer.[2] AES was conceived and implemented by the Examination Division, a user function separate from the Computer Services Division, which maintains the core DP systems for the IRS. It was intended to improve the ability of the Examination Division to cope with the increasing volume and complexity of federal tax returns. In other words, the primary objective of AES was increased productivity. Although AES provided revenue agents with new tools to approach their work, it did not alter the tasks performed since these are heavily constrained by legal and procedural requirements.

As initially implemented, AES consisted of two types of software. The first type included standardized applications which could not be modified by the end user.[3] Revenue agents who used only this type of software would be considered "non-programming end users" (Rockart and Flannery 1983). The second type of software was a commercially available integrated software package including word-processing, spreadsheet, database, and communications. This package contained no tax-related applications, but revenue agents created and shared templates and programs that were useful for auditing. Users of this package would be considered "command level end users" and, in some cases, "programming end users" (Rockart and Flannery 1983).

Although the data collected for this study represent a single job title within a single organization, that job (and the mix of tasks and technology it entails) is not uncharacteristic of white collar, professional work in general. First, revenue agents are not atypical of other computing end users. In the Rockart and Flannery taxonomy, revenue agents represent non-programming, command level, and programming end users. Second, like many salaried professionals in the private sector, revenue agents charge their time directly to specific projects (in this case, audits) and are held strictly accountable for how their time is spent. Like other professionals, the work requires a considerable amount of skill, training, and credentials; the revenue agent position requires a college accounting degree (or equivalent) and approximately 17 percent of all revenue agents are CPAs. Finally, the actual software tools in use are similar to those found on desktops throughout corporate America: spreadsheets, word processing, databases, and an assortment of special purpose, user developed applications.

4.1 Sources of Data

Data from two surveys will be used to examine the relationship between use and productivity. The first is a survey of revenue agents (N = 1110) which was conducted by the IRS in January-February 1988 for the purpose of evaluating the progress of the implementation.[4] The sample was clustered by field audit group (the first level administrative unit of the examination division, consisting of 12 to 14 revenue agents and their group manager). The overall response rate (at the individual level) was estimated to be 85 percent. The survey includes respondents from all seven IRS regions and 48 out of 64 district offices. Although the survey was not designed specifically for this research, it provides reliable indicators for the constructs of interest. Data from this survey will be referred to as "subjective," since the measures of use and productivity are based on self-report.

The second source of data is an independent survey of actual case files (N = 1851) closed during the month of June, 1988, which provides objective measures of use and productivity.[5] This survey was also conducted by the IRS to assess implementation progress. Since audits often take several months, cases closed in June represent roughly the same population of audits that would have been initiated in the first quarter of 1988. Therefore, the time delay between the two surveys is not as substantial as it might first appear. The sample of cases was created completely independently of the first survey and was clustered by revenue agent. There were 418 revenue agents from 21 districts. All seven IRS regions were included in the sample. Of these individuals, 324 were general program agents doing cases comparable to the agents in the first survey. Since these were not the same individuals included in the subjective survey, no within-subjects comparisons between the two data sets are possible. Data from this survey will be referred to as "objective," since the measures are based on the inspection of actual completed audits.

In addition to these survey data, the author has conducted over 100 unstructured interviews of revenue agents and their managers in four different districts over an 18 month period roughly centered on the time of the first survey. These interviews provide an invaluable resource for interpreting the results of the two surveys.

These data will be used for three separate analyses. The first analysis will compare the self-reported use of various system features to the objective data. This comparison demonstrates the importance of obtaining objective measures whenever possible. The second analysis examines the relationship between system use (as reflected by the indicators from the first analysis) and independent measures of efficiency and effectiveness. In terms of the model, this analysis estimates the relationship between use and performance while holding task-technology fit and skills constant. The third analysis will examine the effect of user discretion, training, management, and user characteristics on use, efficiency and effectiveness. This analysis examines other factors that may influence use or performance by affecting the underlying motivational or skill levels of users.

4.2 Operationalizing Use and Productivity

Use is operationalized in two different ways in this paper, as shown in Table 1. For the first two analyses, use is an independent variable operationalized as the collection of specific features used at the discretion of the revenue agent. The objective survey indicates which features were actually used for each case (e.g., database, spreadsheet, word processing), while the subjective survey asked users to indicate which features of the software they use ("Which of the following do you use? Check all that apply."). Since these items are directly comparable, they are used to make the comparison between self-reported and objective use in the first analysis and to predict efficiency and effectiveness in the second analysis. For the third analysis, use is an independent variable, operationalized as overall utilization as indicated by the number of audit tasks automated (e.g., computations, workpapers, document requests), the number of AES features used (e.g., word processing, spreadsheet, workcenter), and the hours of use per week.[6]

In the analysis that follows, productivity will be operationalized in terms of both efficiency and effectiveness, since revenue agents can become more productive by improving their performance along either of these dimensions. Efficiency represents quantity (or how quickly a given task is done), while effectiveness represents quality (or how well a given task is done) (Keen and Scott Morton 1978). Both dimensions are important for evaluating overall productivity changes, because even if the computer does not speed up the audit process, it might make the product qualitatively better.

Efficiency is operationalized in a straightforward manner as the difference in the time required to complete a case when the laptop is used. "Time-on-case" is a basic criteria for productivity in the examination division. Although individual agents are not evaluated on this measure, aggregate measures of time-on-case appear on district and regional level reports as a key indicator of performance. The subjective survey includes two items regarding time on case, which make a reasonably reliable scale (see Table 2 in Section 4.3.3). In the objective survey, the actual hours charged to each case are included in the data. The two measures are comparable because completed audits are the sole work product of revenue agents.

Effectiveness represents the difference in the quality of work done using the computer and is more difficult to operationalize. The items from the subjective survey chosen to measure effectiveness are five point Likert scales, indicating agreement or disagreement with the following statements:

- "The laptop improves the quality of work."
- "The laptop enhances my ability to contribute to my job."
- "The laptop enhances my pride in my work."
- "The laptop allows me to do more."

These items were chosen because they reflect that work is being done better or more thoroughly. To the extent that these items are similar to those used by Robey (1979) and Doll and Torkzadeh (1988), it could be argued that they reflect general user satisfaction as well, but not the more specific construct of "information satisfaction" (Ives, Olsen and Baroudi 1983). The objective survey provides no directly comparable measure of effectiveness, but it does provide a reasonable proxy, which is the yield (dollars assessed per hour of time on case). Given a particular mix of cases, high yield reflects effective use of time by the auditor.[7] It does not capture other dimensions of audit quality, such as freedom from error or clarity of presentation. Although imperfect, these measures of effectiveness augment the efficiency-based productivity measures to give a fuller picture of the overall productivity impact of the AES system.

4.3 Other Variables Used in the Analysis

Several additional variables used in the analysis are also described in Table 1. "Marginal uses" is indicative of user discretion, because it reflects behavior which diverges from strictly efficiency maximizing. The survey item asks "How often do you use the computer for things that would be faster to do by hand?" Allowable responses (and the observed distribution) included "never" (12.5%), "rarely" (16.8%), "sometimes" (52.6%), and "frequently" (18.1%). Clearly, some of these respondents were exercising some criteria other than efficiency for deciding whether to use the computer.

Table 1. Independent and Dependent Variables

Variable	Number of Items	Alpha	Description
Dependent:			
Overall Use (subj)	3	0.78	Tasks, features and hours of use
Efficiency (subj)	2	0.72	Does it take more or less time to complete your work with PC?
Efficiency (obj)	1	n.a.	LN(Actual time to complete audit)
Effective (subj)	4	0.86	Items on quality and effectiveness
Effective (obj)	1	n.a.	LN(Actual dollars per hour)
User Discretion:			
Features (Subj)	6	n.a.	Word Proc, spreadsheet, Database, Workcenter, RAR, and field use
Features (Obj)	6	n.a.	Word Proc, spreadsheet, Database, Workcenter, RAR, and field use
Marginal Uses	1	n.a.	Do you sometimes use the PC for things that would be faster by hand?
Training:			
Instructor	2	0.82	Instructor ability and experience
On-the-job	4	0.74	Quality of on-the-job training
Gap before use	1	n.a.	Gap between training and first use of computer on the job
Management Policy:			
Use Required	4	0.70	I use the computer because I feel I have to, I will be evaluated on it, etc.
Time to learn	1	n.a.	How helpful was management policy regarding time on case?
User Characteristics:			
Prior Exper	1	n.a.	Prior experience with computers
Time with PC	1	n.a.	Time between training and survey
Grade level	1	n.a.	Grade level of respondent
Age	1	n.a.	Age of respondent (7 point scale)

Training is included in the analysis because it is an important predictor of success in end user computing (Cheney, Mann and Amoroso 1986; Fuerst and Cheney 1982) and because it can be a proxy for user skills. Although all users received the same formal training (two weeks of classroom instruction and four to six weeks of on-the-job instruction), some important differences were captured on the survey. "Instructor" measures the perceived experience and training ability of the instructor. "On-the-job" measures the quality of on the job training, as indicated by the number of distinct activities included (full-time instructor, field visits, training cases, workshops). "Gap before use" measures the number of weeks elapsed between training and actual use. Such gaps usually resulted from temporary job assignments, such as answering taxpayer questions during March and early April.

The management policy variables included on the survey were somewhat ad hoc, but reflect important dimensions of the managerial stance towards use of the computer. "Use required" reflects the perception that use would soon become a job requirement.[8] "Time to learn" reflects the perception that management policy allowed agents to put additional hours on cases as long as they were using the computer. High values of this variable reflect situations where the agents were able to take more time to learn.

User experience is also of interest in this analysis as a proxy for skill. "Prior exper" reflects the prior experience of the user with computers, while "Time with PC" measures the time since the user received AES training. "Grade level" captures job tenure and also differences in the kinds of work agents do as they become more senior. Junior agents do large numbers of simple audits, mostly 1040s, while senior agents do fewer, more complex cases. The age of the user was collected on the survey and is also included in this analysis.

4.3.1 Analysis 1: Comparison of Actual and Self-Reported Use

Table 2 compares the actual and self-reported use of various system features. The first column represents actual use by case, while the second column shows actual use by agent. (In this table, an agent is counted as using a feature of AES if they use it on any of the cases in the sample.) It is apparent that users greatly over-estimated their actual use of the system for every feature. Use of the spreadsheet, database and workcenter are particularly overstated. For random samples of this size, even after adjusting for clustering, the confidence interval around the population means reported is only a few percent. Selection bias cannot account for differences of this magnitude.

Table 2. Comparison of Actual and Self-Reported Use of AES Features

Feature of AES	Actual Use by Audit (N=1316) Mean	Actual Use by End User (N=324) Mean	Self-reported by End User (N=1088) Mean
Word Process	71.4%	77.1%	95.3%
Spreadsheet	25.8%	41.9%	82.1%
Database	2.5%	5.4%	30.7%
Micro RAR	66.9%	70.2%	80.3%
Workcenter	6.2%	15.2%	31.9%
Field Use	34.5%	46.5%	74.3%

It is more likely that users responded to the self-report survey based on features they **were taught** to use, **are able** to use, or **would like** to use, rather than features that they **actually do** use. The tendency to overstate use makes sense, because use of the computer was seen as socially desirable; the survey was "looking for" use, so a response bias in that direction is to be expected. This simple comparison does not test any part the model per se, but it does demonstrate the point that objective measures of use can be rather different than self-reported measures.

4.3.2 Analysis 2: Empirical Relation between Use and Productivity

The relationship between the use of various AES features and the efficiency and effectiveness of the users is shown by the regressions in Table 3. These regressions directly estimate the key relationship in the model between use and productivity, while implicitly controlling for skills and fit. The entries in the table are the standardized beta-weights. Since the variables themselves are dummies, the crucial feature of these regression coefficients is their sign. Positive coefficients represent improved productivity.

An inherent problem in a regression like this is that users are self-selecting the kinds of cases to which they apply the technology. If they tend to apply the computer to larger cases, then any efficiency gains could be overwhelmed by the selection effect. If users are applying the technology productively, one would expect to see some positive coefficients in a sample of this size.

In the subjective data, only use of the database is associated with higher efficiency. Most system features have no effect, but the standardized workcenter program has a small negative association with efficiency. Effectiveness, however, appears to be strongly influenced by use of word processing, spreadsheet, database, micro RAR, and use of the computer in the field. Only the workcenter appears to be negatively associated with effectiveness. The conclusion based on the self-report data would be that AES has little effect on efficiency, but it has a substantial positive influence on effectiveness. In other words, users are satisfied with the system even though it does not help them complete their work faster.

The objective data tell a somewhat different story. First of all, none of the features of AES are positively associated with increased efficiency and three of them are negatively associated. Increased effectiveness appears to be associated with the use of word processing and the micro RAR for tax computations. Based on these observations, one might be tempted to conclude that the computer damages efficiency somewhat, but perhaps makes up for it in improved effectiveness. Recall that our measure for effectiveness is dollars per hour. It is unlikely that word processing helps revenue agents identify understated income or overstated deductions; the primary use for word processing in an audit is to write up the audit trail and the

Table 3. Use as a Predictor of Efficiency and Effectiveness

AES Feature	Subjective Data: Efficiency	Subjective Data: Effectiveness	Objective Data[9]: Efficiency	Objective Data[9]: Effectiveness
Word Process	.051	.125***	.000	.167***
Spreadsheet	.036	.105**	-.171***	-.040
Database	.140***	.125***	-.085 **	-.018
Micro RAR	.045	.140***	.008	.118***
Workcenter	.093**	-.070*	.098**	.045
Field Use	-.048	.104**	-.036	.011
Multiple R	.188	.348	.253	.236
Adjusted R^2	.029	.116	.059	.051
DF Regression	6	6	6	6
DF Residual	1003	1003	1309	1309
F-value	6.17***	23.05***	14.92***	12.92***

Significance levels: * = (p < .1) ** = (p < .01) *** = (p < .001) results of the examination. It could be that word processing cuts down on the time to create these documents, thereby increasing yield. This possibility is ruled out by the coefficient on word processing in the efficiency equation; word processing has no effect whatsoever on the time required to complete cases.

What explains the relatively strong association between word processing and effectiveness? The answer can be found in these comments from a senior revenue agent and a first line manager regarding the relative merits of handwritten versus computer generated reports:

> **Revenue Agent**: A 30 to 50 page handwritten, pencil report to collect $50,000 or send someone to jail? That's silly! So a computer output looks more professional, for sure. But on a personal level, it enhances people's self-image, too.
>
> **Group Manager**: An important thing to realize about the PC is that because it produces a better looking product, more legible, more comprehensible, it eases the selling job we have to do on the adjustments. You see, when you're out there doing an audit and it comes time to close the case, what you are really doing is selling your skills and image as a credible, competent examiner. Would you sign an agreement form prepared by someone you didn't trust to do it right? Probably not. That's part of why the PC is so important for agents, because a better looking product is easier to sell.

These comments imply that when an agent gets a case with large adjustments that the taxpayer is likely to object to, the agent will prefer to type up the report. Computer generated reports look "more professional," so they enhance the prestige, credibility, and self-image of the auditor. In effect, the computer has value that is more immediate and relevant to the users than productivity. As a result, the causality between "effectiveness" and use leads in the opposite direction from what the model implies.

The second comment also implies that use of word processing actually makes taxpayers more likely to accept an unfavorable audit result. This proposition was tested on the sample of closed cases and exactly the opposite relationship was found. The frequency of taxpayer "disagreement" of the audit results was 4.0 percent in cases done by hand ($N = 499$) versus 10.5 percent in cases with word processing ($N = 1113$) ($t = -4.35$, $p < .000$). Once again, the causality underlying this relationship needs to be examined. Revenue agents are clearly applying the computer to cases where taxpayers disagree with the results more often than they apply the computer to other cases. It is possible that they are doing so because workpapers for "unagreed" cases are usually longer and require more revisions than for "agreed" cases. Indeed, interviewees often commented that word processing was especially helpful for "unagreed" cases for exactly these reasons. If so, then controlling for taxpayer disagreement could reveal a positive relationship between word processing and efficiency. When taxpayer disagreement was controlled for, the results for efficiency and effectiveness were unchanged in magnitude or sign for any of the AES features. Agents employ AES selectively on different kinds of cases, but when this selectivity is controlled for, there is still no positive influence on productivity. Based on these data, one is forced to conclude that the value of the personal computer to this group of users is primarily symbolic.

4.3.3 Analysis 3: Effects of User Discretion, Training and Management

This analysis uses the self-report data to explore the possible ways in which management can encourage users to exercise their discretion in the best interests of the organization. Table 4 presents beta-weights from three regressions which predict use, efficiency and effectiveness as measured by the subjective data. In terms of the model, the training and experience variables are proxies for user skill. "Marginal tasks" represent a particular dimension of user discretion, the central premise of the model. The variables regarding management policy go beyond the constructs explicitly included in the model and speak to the issue of motivation for use.

With the caveat that these data are all self-reported (and demonstrably over-optimistic), there are some interesting observations that can be made. First, discretionary use of the PC for "marginal tasks" (those that would be faster by hand) contributes to the overall level of use, but predictably decreases efficiency. It could be that users are electing to do things more slowly in order to do them better, but this variable also decreases effectiveness. An identical pattern can be seen in the management variable, "use required." Those agents who believe use will be mandatory use the computer more, but they believe they are both less efficient and less effective as a result. This is in striking contrast to the other management policy variable, "time to learn," which is positively associated with all three dependent variables. These data suggest that it is better policy to **allow** the use of personal computers than to **require** it. Although motivation is outside the scope of the model, it is clear that the motivational structure created by management can influence the productivity impact of a system by influencing how and where it is applied.

As expected, the training variables are all significant, but their pattern of influence is interesting. The quality of the instructor has little effect on use, but seems to boost both measures of productivity, while the quality of on the job training affects use but has little effect on either measure of productivity. This would suggest that different kinds of training may affect outcomes differentially, but the results

Table 4. Regression on Use, Efficiency and Effectiveness

Independent Variable	Use	Efficiency	Effectiveness
		Dependent Variable	
Marginal tasks	.175***	-.163***	-.075*
Management Policy:			
Use Required	.087**	-.183***	-.129***
Time to Learn	.118***	.095**	.181***
Training:			
Instructor	.049*	.104**	.163***
On-the-job	.128***	-.072*	-.017
Gap before use	-.257***	-.038	-.114***
Experience:			
Prior Exper.	.195***	.092*	.164***
Time with PC	.200***	.042	.031
Demographics:			
Grade level	-.066*	-.074*	-.082*
Age	-.018	.056	.046
Multiple R	.526	.344	.416
Adjusted R^2	.269	.109	.164
DF Regression	10	10	10
DF Residual	942	942	938
F Value	36.11***	12.62***	19.61***

Significance levels: * = ($p < .1$) ** = ($p < .01$) *** = ($p < .001$)

here are too weak to draw firm conclusions. The gap between training and actual, however, was a negative influence across the board. Apparently, skills acquired in training must be put to use immediately or they will be lost or at least diminished.

The two measures of user experience provide an interesting contrast. Prior experience is associated with increased use and effectiveness, but only very weakly with increased efficiency. If prior computer experience provided these users with transferable skills, those skills translate only into improved effectiveness, not improved efficiency. Experience with the actual system in question, as indicated by "time with PC," is associated only with increased use. If experience improves skills, then the relationship between skills and performance it is not reflected in these self-reported productivity data. Finally, user grade level is mildly associated with lower levels of use and productivity, while user age per se is not. This would suggest that users with higher seniority are less willing to adopt computers in their work, but that the effect is related more to job tenure than to respondent age.

5. DISCUSSION

The use of personal computers by Revenue Agents is an example of the generous deployment of end user computing technology with unclear results. Users believe they are more effective (although not necessarily more efficient), but this claim is difficult to substantiate. The problem is not that the criteria for efficiency and effectiveness are vague or hard to measure. Clear, objective indicators of use and productivity reveal no positive association with efficiency for any feature of the system. Use of certain features correlates with increased effectiveness, but a closer examination of the data indicates that these voluntary users are applying the tools in ways that suit their needs for self-esteem and professionalism rather than the needs of the Examination Division for higher productivity.

In fairness to the revenue agents who use AES, they were limited to the tools they were given. The limited productivity gains they achieved could be caused by another key element in the model: task-technology fit. Since all revenue agents were performing the same general mix of tasks with identical hardware and software tools, the fit (or lack of fit) between task and technology was controlled for. Although the lack of variance on this dimension precludes a quantitative analysis, the results presented here are generally consistent with a poor overall fit. There was strong agreement among survey respondents and interviewees that the computer was an excellent fit for some tasks (e.g., computation of tax and penalties), but these tasks appear to make up a small portion of the overall workload of a revenue agent. For the remaining tasks, the fit was probably considerably worse. In the summation over tasks, the net effect could easily be negative. Only those agents who were appropriately selective in the use of the computer would achieve an overall productivity gain.

Insufficient user skills and training could also account for the low productivity impact of AES. Since all users received the same training, the variance on this dimension is low and no definitive statements are possible from this data. User experience shows a small positive association with overall productivity, which suggests that there was room for improvement in user skills through better training, and many interviewees expressed a desire for additional training. The small size of the association of productivity and experience implies that even the best users were not markedly better overall than the worst. Once again, this finding is consistent with relatively poor task-technology fit.

If this is so, it bodes ill for the productivity impact of personal computing in other situations as well. These accounting professionals received very widely accepted "personal productivity" tools and several weeks of hands-on training in the classroom and on the job. If spreadsheets and word processing are of limited benefit to these professionals, whose primary work product is written reports filled with tables of numbers, then who are these tools going to benefit?

The model presented here suggests that user discretion is the key to this question. One is tempted to paraphrase a familiar political slogan: "Computers don't save time, people save time." In principle, almost any technology can boost productivity if is applied by a skilled user to the right task. For end user computing to have an overall positive influence on user productivity, it must be applied to the right mix of tasks. For voluntary users, that mix is up to their discretion.

The message to MIS researchers is that system use should not be accepted as a proxy for system success, even for voluntary users. Too many factors can intervene to distort or destroy the assumed relationship between voluntary system use and actual system benefits. Practitioners should also be aware that more use is not necessarily better. Implementation policies that encourage use may also lead to unproductive applications of technology.

6. FUTURE RESEARCH

The results presented here could be improved and extended in a variety of ways. First of all, it would be useful to operationalize and estimate the effects of all of the variables implied in the model: use, task-technology fit, and user skills. For example, it may be that good fit between task and technology could dominate over (and reduce the need for) user skills, but that is an empirical question that should be explored. Likewise, the simplifying assumption which limited motivation to being a predictor of use should be examined. To the extent that many important implementation policies influence use via user motivation, a more practical model would treat motivation explicitly.

Although the model emphasizes discretionary use, the relationship between use and productivity cannot be reliably studied without directly manipulating use as a variable. When use is strictly voluntary, self-selection by users threatens the internal validity of the results. The analysis presented here attempted to correct for this problem by using large sample sizes and controlling for certain known selection criteria. In general, however, an experimental design where use is randomly assigned would be preferred.

Finally, it would be useful to validate the subjective measures of system effectiveness which are becoming increasingly popular (e.g., user information satisfaction) by comparing them to objective measures using a within-subjects design. Such a comparison would require a research context where objective measures are available and the subjective measures are relevant. Although user satisfaction may be the best available indicator in many cases, the mere fact that it is self-reported should be cause for concern. The data presented here suggest that simple yes or no questions regarding use of system features are not accurately answered. Why should questions regarding timeliness, relevance, or ease of use be any better? Although the answers to questions of this kind have been shown to covary in predictable ways through factor analysis (c.f., Doll and Torkzadeh 1988; Ives, Olsen and Baroudi 1983; Bailey and Pearson 1983), they have not been shown to covary with any objective measure of productivity. To the extent that productivity is an important issue and objective measures remain elusive, the validation of these alternative instruments becomes all the more important.

7. REFERENCES

Bailey, J. E., and Pearson, S. W. "Development of a Tool for Measuring and Analyzing Computer User Satisfaction." *Management Science*, Volume 29, Number 5, May 1983, pp. 530-545.

Benson, D. H. "A Field Study of End User Computing: Findings and Issues." *MIS Quarterly*, Volume 7, Number 4, December 1983, pp. 35-45.

Burrell, G., and Morgan, G. *Sociological Paradigms and Organizational Analysis*, London: Heineman, 1979.

Campbell, J., and Pritchard, R. "Human Motivation." In M. Dunnette, Editor, *Handbook of Industrial and Organizational Psychology*, Chicago: Rand, 1976.

Cheney, P. H.; Mann, R. I.; and Amoroso, D. L. "Organizational Factors Affecting the Success of End-User Computing." *Journal of Management Information Systems*, Volume 3, Number 1, Summer 1986, pp. 65-80.

Crawford, A. "Corporate Electronic Mail: A Communication Intensive Application of Information Technology." *MIS Quarterly*, September 1982, pp. 1-13.

Danziger, J. N., and Kraemer, K. L. *People and Computers: The Impacts of Computing on End Users in Organizations*, New York: Columbia University Press, 1986.

Davis, F. D. *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results*. Unpublished Doctoral Dissertation, Massachusetts Institute of Technology, 1985.

DeSanctis, G. "Expectancy Theory as an Explanation of Voluntary Use of a Decision Support System." *Psychological Reports*, Volume 52, February 1983, pp. 247-260.

Desanctis, G., and Gallupe, R. B. "A Foundation for the Study of Group Decision Support Systems." *Management Science*, Volume 33, Number 5, May 1987, pp. 589-609.

Doll, W. J., and Torkzadeh, G. "The Measurement of End-User Computing Satisfaction." *MIS Quarterly*, Volume 12, Number 2, June 1988, pp. 259-274.

Feldman, M. S., and March, J. G. "Information in Organizations as Signal and Symbol." *Administrative Science Quarterly*, Volume 26, Number 2, June 1981, pp. 171-186.

Fudge, W. K., and Lodish, L. M. "Evaluation of the Effectiveness of a Model Based Salesman's Planning System by Field Experimentation." *Interfaces*, Volume 8, Number 2, Part 2, November 1977, pp. 97-106.

Fuerst, W. L., and Cheney, P. H. "Factors Affecting the Perceived Utilization of Computer-Based Decision Support Systems in the Oil Industry." *Decision Sciences*, Volume 13, 1982, pp. 554-569.

Goodhue, D. "IS Attitudes: Towards Theoretical Definition and Measurement Clarity." *Proceedings of the Seventh International Conference on Information Systems*, 1986, pp. 181-194.

Guimaraes, T. "Personal Computing Trends and Problems: An Empirical Study." *MIS Quarterly*, Volume 10, Number 2, June 1986, pp. 179-187.

Ives, B.; Olsen, M.; and Baroudi, S. "The Measurement of User Information Satisfaction." *Communications of the ACM*, Volume 26, Number 10, October 1983, pp. 785-793.

Jensen, M. C., and Meckling, W. H. "Theory of the Firm: Managerial Behavior, Agency Costs, and Ownership Structure." *Journal of Financial Economics*, Volume 3, October 1976, pp. 305-360.

Keen, P. G. W., and Scott-Morton, M. S. *Decision Support Systems: An Organizational Perspective*, Reading, Massachusetts: Addison-Wesley, 1978.

Kling, R., and Gerson, E. M. "The Social Dynamics of Technical Innovation in the Computing World." *Symbolic Interaction*, Volume 1, Number 1, 1977, pp. 134-146.

Lefkovits, H. C. "A Status Report on the Activities of the Codasyl End-User Facilities Committee (EUFC)." *Information and Management*, Number 2, 1979, pp. 137-163.

Loveman, G. "The Productivity of Information Technology Capital." Unpublished Working Paper, Massachusetts Institute of Technology, 1987.

Lucas, H. C., Jr. "Performance and the Use of Information Systems." *Management Science*, Volume 21, Number 8, April 1975, pp. 908-919.

Maish, A. M. "A User's Behavior Towards His MIS." *MIS Quarterly*, Volume 3, Number 1, March 1979, pp. 39-52.

Markus, M. L. "Power, Politics and MIS Implementation." *Communications of the ACM*, Volume 26, Number 7, June 1983, pp. 430-444.

Myer, A. D. "Adapting to Environmental Jolts." *Administrative Science Quarterly*, Volume 27, Number 2, December 1982, pp. 515-537.

Peters, L. H. "Cognitive Models of Motivation, Expectancy Theory, and Effort: An Analysis and Empirical Test." *Organizational Behavior and Human Performance*, Volume 20, Number 1, October 1977, pp. 129-148.

Pettigrew, A. M. *The Politics of Organizational Decision Making*, Tavistock, London, 1973.

Robey, D. "User Attitudes and Management Information Systems Use." *Academy of Management Journal*, Volume 22, Number 3, 1979, pp. 527-538.

Robey, D., and Markus, M. L. "Rituals in Information System Design." *MIS Quarterly*, Volume 8, Number 1, March 1984, pp. 5-15.

Rockart, J. F., and Flannery, L. S. "The Management of End-User Computing." *Communications of the ACM*, Volume 26, Number 10, October 1983, pp. 776-784.

Sutherland, D. J. "Attitudes and Management Behavior of Senior Managers with Respect to Information Technology." Unpublished M.S. Thesis, Massachusetts Institute of Technology, 1988.

Trice, A. W., and Treacy, M.E. "Utilization as a Dependent Variable in MIS Research." *Database*, Fall/Winter 1988, pp. 33-41.

Vroom, V. *Work and Motivation*, New York: Wiley, 1964.

8. ENDNOTES

1. Personal computing is a type of end user computing which is distinguished primarily by the single-user hardware platform, typically a micro computer (Guimaraes 1986). The implications of this difference are unclear. Rockart and Flannery (1983) note that personal computing generally involves the same managerial issues as end user computing on other hardware platforms. As a result, personal computing is often lumped together with other forms of end user computing (e.g., Doll and Torkzadeh 1988). However, Benson (1983) found striking differences between the kinds of applications run on mainframe versus micro computers, with mainframe users predominantly performing "data capture" and micro computer users predominantly performing "analysis." These differences may be a reflection of the historical situation of Benson's research, since micro computers were a relatively new phenomenon in early 1983. As micro computers become more powerful and more integrated into corporate data resources via networking, the distinctions between personal computing and other forms of end user computing may vanish. The distinction is made here only because the data to be presented are from users of micro computers.

2. The laptop portables are only the first of several phases. The complete system is planned to include networks of mini-computers and mainframes on a national basis, on-line access to taxpayer records, on-line legal research, and artificial intelligence for classification and selection of returns.

3. The two main examples of this class of tools were the 1040 Workcenter (a comprehensive tool for auditing 1040 returns and all of the associated schedules) and Micro RAR (a program which computes tax, interest and penalties, but does not otherwise assist in the audit process).

4. This survey was designed and administered by Price Waterhouse under contract to the IRS approximately eighteen months after implementation started. The sample also included an additional 590 group managers whose responses are not analyzed in this paper.

5. Case files containing taxpayer information are strictly confidential. Data for this survey was collected by authorized IRS personnel; no confidential information of any kind was transmitted to the author. Of the 1,851 cases in the sample, only 1,316 were "general program" cases. The remainder were specialty cases, including excise, estate and gift, etc., and were excluded from the analysis to insure comparability between the two samples.

6. The Cronbach alpha for this scale is shown in Table 1. The construct validity of the scales used here was checked using a factor analysis and by examining item correlations.

7. At first glance, yield would appear to be more related to efficiency than effectiveness. That would be true if the objective of the examination division was to maximize the dollars assessed. Popular opinion notwithstanding, the operational objective of the examination division is to examine a certain number of returns of various types. Given an inventory of returns to audit, yield is a measure of the effectiveness of the auditor in deciding which specific tax issues to pursue.

8. In some locations, management made a concerted effort to induce agents to use the computer by suggesting (without really knowing for certain) that use of the computer would become mandatory within eighteen months. With this in mind, agents felt obliged to begin using the computer immediately.

9. The distribution of these dependent variables were transformed by natural logarithm to be nearly normally distributed. Also, outliers were checked and removed to improve the estimates of the coefficients. The sample used in these regressions reflects the full range of general program audits, from individuals to large corporations. Regressions on sub-samples comprising each of 16 separate audit classes (e.g., 1040, $10K to $25K) yielded the same basic results.