The Effects of Computer Monitoring on Secretarial Performance and Stress

Dennis F. Galletta, Ph.D.
Associate Professor of MIS
Katz Graduate School of Business
University of Pittsburgh
galletta@vms.cis.pitt.edu

Terry L. Huston, Ph.D.
Assistant Professor of MIS
School of Business and Administration
Saint Louis University
hustontl@slu.edu

Special thanks to the Saint Louis University School of Business summer research grant for aiding in this work.

Copyright 1998

Please do not quote or copy or paraphrase in any way without the permission of Terry L. Huston or Dennis F. Galletta. Thank you.
ABSTRACT

Computerized performance monitoring systems (CPMS) are pervasive, and there is speculation, as well as survey and case study evidence in the literature, that they might cause ill effects on performance, and might increase worker stress. This laboratory study investigated possible impacts of CPMS on physiological and perceived stress, as well as on the performance of the subjects.

The study involved 35 [?] secretaries. [need to modify this according to new results: Results indicated that the physiological measures did not seem to differ when comparing the computer-monitored group with the control group. An exploratory human-monitored group also did not seem to produce any striking previous survey findings, and lower state anxiety was found where CPMS might be thought of as appropriate than when they might be considered to be inappropriate. Work quality appeared to suffer as a result of using a computerized monitor, and the speed of completion increased sharply. Efficiency appeared to increase because the amount of time spent decreased much more quickly than did overall quality.] On the other hand, if the highest quality is necessary, then a monitor might not be desirable. Future research should focus on replication with a larger pool of subjects, using other tasks, employing other physiological measures, and considering a longer-term time frame.

Key Words: Computerized performance monitoring, experiment, performance, stress
I. INTRODUCTION

A great deal of recent attention has been paid to the practice of using technology to monitor the performance of employees while their tasks are in process. Among those who have a great deal of interest in this subject are employers who desire to gain greater task control, as well as employees who would prefer to retain their task autonomy and personal privacy. In addition, organized labor and governmental agencies have examined the implications of computerized performance monitoring systems (CPMS), and have invested resources in gathering and communicating information about these systems.

Researchers to date have most commonly made use of survey and case study methods for gathering empirical data about CPMS. The results so far suggest that such systems could indeed have effects on at least self-report outcomes. However, the results are in some cases contradictory, perhaps because of differences in procedures and measures used by different investigators relying on a variety of reference disciplines on which to base their work. Therefore, there appears to be a need for a replicable study in the experimental domain, which addresses a limited set of effects of CPMS on a specific set of workers.

This paper describes an experiment that made use of three typing tasks in conjunction with questionnaires, time measures, double-blind judgments of performance, and physiological measures, in an attempt to uncover any highly specific effects of CPMS. This study tapped literature from a variety of sources in developing the tasks and measures. The next section will describe the important findings revealed by investigating this literature. Subsequent sections will describe the experimental method, results, and limitations and conclusions of the study.

II. Review of the Literature

The literature raises concerns about several interrelated issues associated with CPMS. In general, there are concerns that CPMS can raise stress in workers. There are fears that this stress has adverse effects on physical health, attitudes, and performance. This section reviews literature from several disciplines in addressing these problems.
General Background

The use of CPMS is quickly becoming common in many work settings (George, 1996; Laabs, 1992; Collins, 1987; Foegen, 1987; Westin, 1987; Harz, 1985). Although there have been outrageous claims that over 80-90% of clerical employees are subjected to electronic performance monitoring, Westin's (1987) more careful estimate of the actual extent to which these employees are being monitored and evaluated for pay, promotion, or discipline puts the figure at between 20 and 35%. More recently Aiello (1993) indicates that this figure may be as high as 40%. Not only are clerical workers monitored, but also technical and professional employees (Tinkham and Kleiner, 1993; Brandon and Inman, 1992; Grant & Higgins, 1989; Garson, 1988) in some cases. Given that the combination and trend of recent estimates (Reynolds, 1996; DeTienne, 1994; Nussbaum & duRivage, 1986; Serrin, 1986; Harz, 1985) would lead one to estimate that there are as many as 30 million workers who are monitored, the study of CPMS appears to be warranted at this time.

Monitors can detect whether or not terminals are on or active; count keystrokes or completed transactions; count errors; warn users during periods of declining speed; and even log break periods (DeTienne and Alder, 1995; Tinkham and Kleiner, 1993; Laabs, 1992; Grant & Higgins, 1989; Harz, 1985). Although there is some controversy in the use of such monitors (DeTienne, 1995; Picard, 1994; Anderson, 1993; Laabs, 1992; Chalykoff and Kochan, 1989; Grant, 1989; Pituero, 1989; Grant et al., 1988), the literature appears to be oriented generally toward finding any effects from their use that might exist.

Many studies have linked CPMS to several ill effects, like: decreased quality of work life (Garson, 1988); degraded customer service (Grant et al., 1988); eroded satisfaction (Irving et al., 1986); and increased stress (Aiello and Kolb, 1995; DeTienne, 1994; Smith et al., 1990; Garson, 1988; Irving et al., 1986). Work by Grant (1989) has greatly increased the depth of the CPMS literature. Her results indicate that monitors are not unidimensional constructs, and monitoring a large number of tasks would reduce the acceptability of CPMS. Further, it is important to establish the credibility of the computer as an appropriate measurement device, or the monitor's acceptability will diminish. Grant and Higgins (1991) extended this to find that monitoring may not increase production and, even if it does, it need not reduce the importance of service. They also found that
various monitor features can be altered to change the impact and that the credibility of the computer is a factor in the monitor’s impact.

In another study, Grant and Higgins (1989) imply that four dimensions must be considered in assessing the pervasiveness of monitors: the object to be monitored, the period of monitoring, the recipient of the data, and the tasks that are monitored. Least pervasive CPMS involve the business unit as the monitored object, regular and infrequent monitoring, the employee alone as the recipient, and outcome rather than process-based monitoring. Most pervasive CPMS involve the individual employee, constant and immediate monitoring, public display of results, and monitoring of tasks while in process. Reported acceptability of such monitors seemed to be increased by the first two dimensions, but decreased by the last two dimensions. Overall, however, the more pervasive the monitoring system, the more important it was perceived by employees for evaluation purposes.

There is some indication that monitoring may have some positive effects if performed for feedback purposes to help the employee improve performance although in many cases there is no such feedback provided to the employee. Griffith (1993a) extols the virtues of the responsible use of a computer monitoring system to provide feedback data to the workforce rather than merely to gather social information (e.g., time taken for bathroom breaks). Collecting data that employees believe will help them can make monitoring a more effective tool.

Chalykoff and Kochan (1989) found that the negative impacts of computer monitoring can be mitigated by attention to feedback/performance appraisal processes. In this study, job satisfaction and employee turnover intention were affected by the employees’ affective responses to monitoring and by their prior beliefs or dispositions about monitoring.

Another way to use monitoring effectively is reported by DeTienne and Alder (1995). Electronic monitoring can be used to help train new employees. Consistent with theories of learning, prompt feedback is essential to effective training and electronic monitoring can provide this. An extension of this type of monitoring would be to use it to give employees immediate positive reinforcement and recognition.

Some managerial implications for the use of monitoring stem from the findings that employees react more positively to monitoring when they see it as beneficial to them and they
believe it is being used fairly (DeTienne 1994). Although monitoring is frequently used to increase the work pace and prod the slower workers, managers should be tolerant of diverse working styles. Employees even report nothing but positive feelings about being monitored (Laabs 1992). Once again, this result arises from the immediate use of the monitoring session for positive reinforcement.

Even though the effects of monitoring on stress has been a common concern of practitioners as well as researchers, there is also a concern as to the effect monitoring has on the performance of the individual. One of the primary reasons computer monitoring is utilized in many cases is to enhance the productivity of the employee to the benefit of the employer. Are the effects on performance favorable? Also, how do the effects of computer monitoring compare to that of the more traditional human (supervisor) monitoring when measuring performance?

Griffith (1993b) found that the performance of those in the computer monitored condition was less than that of those in the human monitored condition, although the results were not significant. In this case it was assumed that electronic media are more uncertain regarding evaluation because it is not clear when or how the worker is being monitored. This would not be a case where the computer monitoring is being used for feedback purposes. Aiello and Svec (1993) found that human and computer monitored subjects performed much more poorly than those in a ‘low-monitored’ condition (those who were told that their work would not be monitored).

Human monitoring aside, what effect does the worker’s skill level have on their ability to perform under monitored conditions? Aiello and Kolb (1995) found that performance depended on whether the worker was highly skilled or not. In their study, highly skilled computer-monitored subjects outperformed their nonmonitored counterparts while just the opposite was found in low-skilled participants. Nebeker and Tatum (1993) had also found an effect on performance depending on ability. Subjects who were monitored and given feedback had a higher rate of production but suffered from poorer quality of work, reduced satisfaction, and increased stress than those who were not aware of any monitoring mechanism.

The implications of much of the research would seem to indicate that managers should carefully consider the way in which computer monitoring is used (DeTienne 1994). Recent academic research indicates that computer monitoring is a malleable technology (George 1996).
and that a lot of its bad name comes from the punitive way in which it is used. Management has a key role to play in the design of monitoring systems that employees will tolerate and approve of rather than merely increasing pressure to produce, thereby potentially affecting employee work quality and health.

All of these studies have provided an important basis for further research in CPMS. The self-report measures, observations, and few experiments present compelling evidence that monitors might have strong impacts on workers. There is a need, however, for additional experimental data that can provide further evidence of the effects of CPMS (DeTienne 1994). DeTienne further suggests that future research is needed to examine various conditions under which monitoring is productive or counterproductive and more stressful or less stressful, especially the relationship between monitor-induced stress and performance, which has occasionally been found to be a nonlinear relationship. This study focuses on stress-related outcomes and performance effects that might be associated with highly pervasive systems, especially when the monitoring is not done for positive feedback purposes.

**Stress and Stress Outcomes**

There are several definitions of stress, but one useful definition is provided by Gowler and Legge (1975), where stress is said to be an imbalance a person perceives between demand and capacity to respond, as long as failure to respond carries with it significant undesirable consequences. Selye's (1956) well known General Adaptation Syndrome outlines three stages of stress. The first stage is an alarm reaction, where an individual's defense mechanisms come into play. The second is resistance, where the individual attempts to return to a state of equilibrium. If the stressors continue, the third stage occurs, where the individual's adaptive mechanisms shut down in exhaustion. Selye notes physiological reactions like increased heart rate, blood pressure, and respiration; psychological reactions like fear, anxiety, and tension; and coping behaviors like information-gathering, denial, or drug abuse.

Schuler (1980) reported that stress in U.S. industry costs approximately $45 million per year. These costs result from absenteeism, turnover, accidents, reduced performance, lack of concern, etc. Similar evidence has been found by other researchers (for example, DeTienne and Alder, 1995; Brandon and Inman, 1992; Baron, 1983; Sailer et al., 1982; Van Sell et al., 1981;
Gupta & Beehr, 1979; Lyons, 1971). Since Selye's (1956) theory is that stressors are additive, and Schuler's cost estimate predates the bulk of CPMS, the findings that CPMS can increase stress (Aiello and Kolb, 1995; DeTienne, 1994; Smith et al., 1990; Garson, 1988; Irving et al., 1986) are of particular concern.

Since many of the stress outcomes identified by Selye are measurable, we now turn our attention to the measures that are available for use in collecting physiological evidence of its presence.

**Physiological Stress Measurement**

Psychophysiological measures have been employed for many years in behavioral assessment. Such measures often include heart rate, skin temperature, skin resistance, and cortisol secretion (Jennings, 1990).

It is generally accepted that an increase in stress causes a subsequent increase in heart rate (for example, Jennings et al., 1970; Craig & Woods, 1969; Deane, 1961). These effects are evident even for mental problem solving exercises requiring only internal concentration (Kaiser & Sandman, 1975; Tursky et al., 1970; Kahneman et al., 1969). The fact that some studies showed decreases in heart rate from stress (for example, Fenz & Jones, 1974; Notterman et al., 1952) was explained by Deane (1961), who asserted that heart rate acceleration accompanies anxiety or dread and that deceleration accompanies fear; fear often occurs close to the aversive event. The study by Fenz and Jones showed that poor parachutists exhibited higher heart rates and less deceleration just before the jump than did their experienced counterparts. Our assumption is that computer monitoring is less likely to cause (specific) fear than (more general) anxiety. Several areas on the body are usable for measuring a subject's pulse, including fingers and earlobes. Earlobe measurement is afforded by detecting changes in translucence of the earlobe caused by a surge of blood into the area accompanying each heartbeat.

Assessing stress by measuring skin temperature has somewhat less reliability than measuring heart rate, but taken along with heart rate, produces a better indication of stress than either measurement by itself. It has been shown that the response pattern is very stable for the same subject when studying the effects of stress on configurations of two or more measured variables versus the variables taken separately (Lacey & Lacey, 1958; Lacey et al., 1953; Lacey & Van Lehn,
1952; Lacey, 1950). Thermistors placed on extremities (usually fingers or hands) register the effects of peripheral vasoconstriction. Vasoconstriction causes a decrease of blood flow, and a corresponding decrease in skin temperature in the extremities (Fuller, 1977). Ambient temperature control or measurement (for use as a covariate) is very important in studies using skin temperature as a dependent variable.

Changes in skin resistance can result from secretion of sweat glands, and are most commonly measured by a psychogalvanometer or by a polygraph (Jennings, 1990; Fowles, 1986). Probes are placed on a foot or palm of the hand. A small amount of current is then applied to detect any increase in conductivity due to the moisture. It is also possible to use a microscope to determine the saturation of an absorbent paper placed on a hand or foot (Jennings, 1990).

Cortisol is secreted by the body in times of stress, and is detectable in urine, blood, and saliva. Salivary cortisol is commonly assayed by using a dental wad (using absorbent material placed between the gum and cheek of each subject) (Klaniecki, 1990). A baseline measurement must be taken along with other measurements at important points during an experiment. The moisture from the wad must be frozen, thawed, and centrifuged, to separate the precipitates. Assay of cortisol is an expensive procedure: costing approximately $20 per sample tested. However, this test is quite reliable.

**Synthesis**

The oft-seen general assertion that computer monitoring contributes to stress and lower performance in office workers can be tested to some extent in the laboratory, given the availability of several stress measures. Two hypotheses will be examined in this study:

**H1:** Stress levels of computer-monitored subjects will exceed stress levels of non-monitored subjects.

**H2:** Performance of non-monitored subjects will exceed performance of computer-monitored subjects.
The following section describes an experiment that employed a high-pervasiveness monitor, physiological stress measures, and performance measures.

III. Method

An experiment was performed to examine outcomes of using performance monitoring on administrative assistant subjects asked to perform typing tasks. We decided to employ the widest variety of conditions, tasks, and measures that were feasible under the circumstances of the field setting. In general, subjects were randomly assigned to one of three monitoring types, three tasks were performed by each subject, and several outcome measures were examined.

Subjects, Incentives, and Groups

Thirty-five professional administrative assistants employed by the schools of business of two universities volunteered to participate in the experiment. Subjects were paid $5 for participating in the study, and were promised a chance to win a $15 prize for producing the "best and fastest" work in each of the two non-practice tasks. All subjects were female.

Subjects were in general (see below) randomly assigned to one of three groups by using a random number generator. The first group was the control group, which involved no monitoring of the tasks in process. Subjects assigned to the second (the computer-monitored) group were told that all of their keystrokes would be captured for use by judges later. Subjects in the third (the human-monitored) group were observed by an experimenter who sat next to them with a clipboard while they performed their tasks. The human-monitoring group was included for exploratory purposes, to serve as a standard for comparing the other groups. No hypotheses have been presented for this group, although it is expected that even greater stress, and worse performance, will be evidenced by this group. Interestingly, Griffith hypothesized that computer-monitored individuals would have more stress and lower performance than human-monitored individuals.

Tasks

Subjects were asked to type three paragraphs. The first was a practice paragraph that covered an administrative assistant policy statement of a fictitious organization, that was to be typed exactly as it appeared. The second was a one-paragraph story that described a manager who
violated several of those policies, also to be typed exactly as it appeared. The third was a memo that was to be composed from the administrative assistant’s point of view. This memo was to remind the manager (the administrative assistant’s boss) of broken rules, how he\(^1\) had violated them, and what he should do to avoid such problems in the future.

The practice task was employed to collect baseline measures from the subjects. The story task enabled the assessment of monitoring effects on highly routine tasks, while the memo task enabled the assessment of monitoring effects on more creative, unstructured tasks. Creative tasks are likely to suffer from external monitors if they serve as extrinsic, rather than intrinsic motivators (Amabile, et al., 1986; Amabile, 1985).\(^2\)

Each subject used the character-based word processing package she normally used every day on her IBM or IBM-compatible personal computer. The paragraphs in the practice and the story tasks were printed using a 12-point, proportionally-spaced, Helvetica-like font, so that the spacing on the screen would not match the spacing on the paper for any subject.

Measures

Measures are summarized in Table 1. They were subdivided into physiological measures, self-report measures, and performance measures. Room temperature was also measured before each subject began the practice task.

<table>
<thead>
<tr>
<th>Physiological Measures</th>
<th>Self-report Measures</th>
<th>Performance Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pulse rate</td>
<td>1. State anxiety</td>
<td>1. Speed</td>
</tr>
<tr>
<td>2. Skin temperature</td>
<td>2. Demographics</td>
<td>2. Accuracy</td>
</tr>
<tr>
<td></td>
<td>3. Other perceptions</td>
<td>3. Memo quality</td>
</tr>
</tbody>
</table>

Heart rate was measured using an Invacare FT4200 pulse monitor, with a clip attached to each subject's left earlobe. Heart rate was recorded every ten seconds on a logging sheet. Skin temperature was measured using a Yellow Springs Instrument model 790TA telethermometer, a

\(^1\) The researchers used the male gender for the manager after deciding to avoid any possible confusion in the use of the pronouns “he” and “she” in the text of the paragraphs used. The female gender was used for the administrative assistant in the paragraphs to match the gender of the subjects.

\(^2\) Although prizes and payments in this experiment might be seen as a hindrance to intrinsic motivation and hence destructive to the creative task, the payment scheme was the same for all subjects in all groups, and it was assumed that additional extrinsic motivation caused by the monitor could still affect creativity.
729 skin surface probe, and a YSI 4009 heat-reflecting shield. The probe was attached to each subject's left cheek, and skin temperature was recorded to the nearest .25 degree Fahrenheit every ten seconds. Extension lines allowed the meters to be behind the partition at all times.

It was decided that skin conductance and cortisol secretions would not be appropriate measures for our usage. The subjects needed to have unencumbered hands during their typing tasks, and it was considered too obtrusive to ask these subjects to remove all footwear and stockings for attaching a foot probe. Likewise, although the experimenters obtained an offer for a limited number of cortisol assays at no cost, dental wads were also considered too obtrusive for these subjects, as would be requests for samples of blood or urine. Such measures could be entirely appropriate for other experimental settings, however, and should not be permanently discarded.

Self-report measures were limited to responses on the Spielberger State-Trait Anxiety Index (STAI) form Y-1\(^3\), on a short demographic instrument, and on a set of Likert questions that covered all of the salient features of the experiment. That last set of questions covered perceived task difficulty, task enjoyment levels, and awareness of the test equipment and other features of the experiment.

Performance measures were speed, where all three tasks were measured by a stopwatch; accuracy, where the practice and story tasks were thoroughly examined and reviewed for typing errors by a panel, and quality, where each original memo was judged, in summary, on whether the instructions were followed adequately, whether the text seemed to be reasonable in light of its purpose, and whether it was typed properly. The memo text generated by each subject was assessed by a panel of two judges against the detailed evaluation criteria. The individual scores for each attribute were summed to derive a total score. The final score for each memo was assigned by averaging the judges' total scores. All accuracy and quality judgments were made in a "double-blind" manner, where the subjects' documents were given random codes to disguise them (and especially their assigned experimental groups).

Procedure

Each subject was seated at a swivel chair and given a general set of instructions to

---

\(^3\) Consulting Psychologists Press, 577 College Ave., Palo Alto, CA 94306.
read. The experimenter reviewed the instructions and asked each subject to sign an informed consent form. After cleaning the equipment with an alcohol swab, the clip was attached to the subject's earlobe and the probe to the subject's cheek as outlined above. The subject was then asked to complete a preliminary STAI form. After completing the STAI instrument, the subject was then asked to swivel the chair to face the computer, and was asked to indicate when she was ready to begin typing. By following this procedure, there was a five-minute adjustment period before taking baseline readings as prescribed by several researchers (Lichstein et al., 1981; Sallis & Lichstein, 1979; Myers & Craighead, 1978) was ensured.

For the practice task, the experimenter immediately went behind the partition. When the subject gave the signal, a second experimenter behind the partition started the stopwatch and began recording pulse and skin temperature readings every ten seconds. When each subject was done, she gave another signal, and the clock was stopped. The experimenter then returned to the subject and asked her to complete another STAI form. He then oriented her to the next task (the story) and allowed the subject to ask any questions.

For the control group, the experimenter then went behind the partition and waited for the subject to indicate when the stopwatch should be started. In the computer-monitored group, the experimenter then explained that a program would capture all keystrokes, including backspaces and arrow movements, for use by the judges in determining the winners. He then pressed a sequence of keys, explaining that the program was now activated for the actual task, and retreated behind the partition. In the human-monitored group, the experimenter sat next to, and slightly behind, the subject so that his presence would be obvious.

Following completion of this task, the experimenter again asked the subject to complete an STAI form, and then oriented her to the memo task. Again the experimenter followed the same procedure as in the story task, took his appropriate position, and waited for the subject to give the signals. Subjects were allowed to study the practice paragraph and the story paragraph for a short time before beginning (and starting the clock), since the memo was intended to integrate the two. After the memo was typed, the subject was asked to complete the final STAI form, was paid the $5, and was asked to sign a receipt and to refrain from discussing the experiment with others.
Prizes were awarded on the basis of the sum of the two standardized z-scores computed from the raw speed score and the raw accuracy or quality score. This served to weight speed and accuracy the same, as promised. One prize was awarded in each of the three experimental conditions, and for each of the two non-practice tasks.

IV. Results

Because the sample size was small, only preliminary findings can be presented. The intention of this experiment was to discover whether or not there was sufficient variability in the physiological measures to warrant their capture; approximately 4,400 physiological observations were recorded, entered, plotted, and averaged. Elimination of one of the physiological measures would cut this requirement in half. Indeed, in future experiments skin temperature readings should probably not be taken. All subjects in all groups appeared to exhibit an identical pattern of skin temperature: a nearly one-degree rise over all three tasks. Therefore, skin temperature will be dropped from the analysis, to save space.

Scores for each of the variables in each group will be shown, in raw form as well as adjusted for baseline measures.

Pulse Scores

Table 2 depicts mean pulse rates for subjects in each group. It appears that pulse rates slowed over time, during execution of the three tasks. The pulse rates for the computer monitored subjects slowed in a similar pattern to those of the control group subjects. Interestingly, the human-monitored subjects did not seem to exhibit the same slowing until performing the creative task. If this is a non-random difference, it is possible that the presence of a person staring over their shoulder for the first task eligible for a prize was slightly more disturbing, and that this disturbance wore off during the third task. Overall Hypothesis 1 does not seem to be supported by either physiological measure.
Table 2 - Mean Pulse Measures

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Computer-Monitored Group</th>
<th>Human-Monitored Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice task</td>
<td>81.7</td>
<td>88.2</td>
<td>87.0</td>
</tr>
<tr>
<td>Story</td>
<td>79.9</td>
<td>85.1</td>
<td>86.4</td>
</tr>
<tr>
<td>Memo</td>
<td>77.2</td>
<td>83.7</td>
<td>83.0</td>
</tr>
<tr>
<td>Percent of base</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>97.8%</td>
<td>96.5%</td>
<td>99.3%</td>
</tr>
<tr>
<td>Memo</td>
<td>94.5%</td>
<td>94.9%</td>
<td>95.4%</td>
</tr>
</tbody>
</table>

Perceived Stress

Table 3 depicts mean STAI scores for subjects in each group. The relative scores appear to behave in a “crossover” pattern when comparing the control group subjects with computer-monitored subjects. Surprisingly, the human-monitored subjects appeared to behave more like the control group subjects. Only speculation can provide an explanation for this phenomenon. It is possible that computer-monitored subjects regarded the monitor as perfectly appropriate for the story (routine) task, but as entirely inappropriate for the memo (creative) task. In addition, subjects would regard a human monitor as appropriate for the creative task as well, but not for the routine task. This speculation is in agreement with the findings of Grant and Higgins (1989), where appropriateness of the monitor is of high importance to any evaluation of the use of a monitor.

Table 3 - Mean STAI Scores

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Computer-Monitored Group</th>
<th>Human-Monitored Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice task</td>
<td>35.3</td>
<td>38.2</td>
<td>41.7</td>
</tr>
<tr>
<td>Story</td>
<td>35.7</td>
<td>36.3</td>
<td>44.3</td>
</tr>
<tr>
<td>Memo</td>
<td>33.0</td>
<td>38.3</td>
<td>37.5</td>
</tr>
<tr>
<td>Percent of base</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>101.1%</td>
<td>95.0%</td>
<td>106.2%</td>
</tr>
<tr>
<td>Memo</td>
<td>93.5%</td>
<td>100.3%</td>
<td>89.9%</td>
</tr>
</tbody>
</table>
Time Taken

Table 4 depicts the amount of time taken by subjects in each group. There appear to be pronounced differences between subjects in the control group, versus computer-monitored groups. After correcting for the typing speed in the practice task, it appears that the computer-monitored subjects worked faster than did the control group subjects. The effect is small for the story task, but is quite pronounced for the memo task. The computer monitor appears to have some effect on the amount of time subjects are willing to spend in completing their tasks. The human-monitoring subjects appeared to be even more inclined to hurry in performing the creative task.

Table 4 - Mean Time Taken

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Computer-Monitored Group</th>
<th>Human-Monitored Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice task</td>
<td>5.4</td>
<td>5.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Story</td>
<td>4.9</td>
<td>4.7</td>
<td>4.8</td>
</tr>
<tr>
<td>Memo</td>
<td>15.1</td>
<td>10.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Percent of base</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>90.7%</td>
<td>83.9%</td>
<td>92.3%</td>
</tr>
<tr>
<td>Memo</td>
<td>279.6%</td>
<td>191.1%</td>
<td>123.1%</td>
</tr>
</tbody>
</table>

Errors Made

Errors are an important consideration when examining typing speed in a routine task. Table 5 presents the mean number of errors made by subjects in each group when typing the story. It appears that the computer-monitored group suffered from slightly more errors in the story as a percent of errors made in the practice task than the control group. Interestingly, the human-monitored group again behaved in a manner more similar to the control group than to the computer-monitored group. The differences are slight enough, however, to attribute this to chance.
Table 5 - Mean Errors Made

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Computer-Monitored Group</th>
<th>Human-Monitored Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice task</td>
<td>13.0</td>
<td>13.8</td>
<td>22.3</td>
</tr>
<tr>
<td>Story</td>
<td>6.5</td>
<td>7.3</td>
<td>11.3</td>
</tr>
<tr>
<td>Percent of base</td>
<td>50.0%</td>
<td>52.9%</td>
<td>50.7%</td>
</tr>
</tbody>
</table>

Quality of Original Text Generated

Table 6 presents the mean quality points in each group earned for creating the original memo. The control group appeared to benefit from the extra time they spent in performing this creative task, and the computer-monitored group suffered somewhat from their shorter time frame. The quality seemed to suffer the most striking degradation in the human-monitored group, who also spent the least amount of time.

Table 6 - Mean Quality Scores

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Computer-Monitored Group</th>
<th>Human-Monitored Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Story</td>
<td>58.2</td>
<td>54.4</td>
<td>41.1</td>
</tr>
</tbody>
</table>

Time-based Units

What might be more important than isolated comparisons of time taken or quality of output could be a combination of the two measures. If evaluation of a secretarial task is seen as one of attempting to maximize efficiency (i.e., having the best outcome per unit of time spent), then per-minute measures might be very relevant to examine. Table 7 presents time-based scores. It appears from this data that computer-monitored subjects performed better than control group subjects, in both routine typing as well as memo creation. This is contrary to the second hypothesis. Even more striking is the markedly higher performance in the memo task by human-monitored subjects than any others. It is obviously due to the markedly lower amount of time spent by human-monitored subjects, but it is difficult to explain how quality remained fairly high for them.
Apparently, the extra time spent by subjects on this task has a relatively low payoff after some threshold.

This result might be considered to be irrelevant by some, since there might be a minimum quality level not achievable in the extremely short amount of time spent by subjects in the human-monitored group. Therefore, these results should be interpreted with appropriate caution.

**Table 7 - Error and Quality Scores as a Function of Time Taken**

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Computer-Monitored Group</th>
<th>Human-Monitored Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error scores per minute</td>
<td>2.7</td>
<td>2.7</td>
<td>4.2</td>
</tr>
<tr>
<td>Practice task</td>
<td>2.7</td>
<td></td>
<td>4.2</td>
</tr>
<tr>
<td>Story, raw</td>
<td>1.4</td>
<td>1.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Story, % of baseline</td>
<td>54.0%</td>
<td>62.0%</td>
<td>56.6%</td>
</tr>
<tr>
<td>Units of quality per minute</td>
<td>4.7</td>
<td>6.7</td>
<td>8.5</td>
</tr>
<tr>
<td>Memo</td>
<td>4.7</td>
<td>6.7</td>
<td>8.5</td>
</tr>
</tbody>
</table>

**Summary**

In sum, the physiological measures did not seem to provide evidence of monitor-induced stress. A perceptual measure appeared to support the findings of previous work by Grant and Higgins (1989). Striking performance differences surfaced, where the computer-monitored group appeared to work much more efficiently, but suffered a loss in overall quality as a result. The next section of this paper will describe limitations, and list some final conclusions from this study.

**VI. Limitations and Conclusions**

This preliminary study supported the assertion that CPMS can have effects on performance and stress of individuals subjected to these systems. However, there are some important limitations that should be made explicit.

**Limitations**

First and foremost, the sample size was small--there were simply not enough degrees of freedom to allow formal hypothesis testing. This was because of the enormous number of measures per person, and future studies can make use of automated data collection instruments to facilitate the use of larger samples. The most important covariate we would use in a
MANCOVA test would be each subject's self-reported typing speed, and this was not possible here. However, the use of the practice task as a baseline measure helped in controlling for this variable, and the mean reported speeds were very similar in each cell.

Secondly, the scripts and task were not taken from previous work, and might benefit from refinement. Some subjects doubted their chances of winning. Some were stressed by the topic of the paragraphs because the situations they faced on a daily basis were unexpectedly very similar to the exaggerated content.

Also, very promising physiological measures (galvanic skin response and salivary cortisol) were not taken because of their obtrusive nature. Another research setting should be sought that would support such measures.

Finally, only short-term phenomena were measured. It is unknown if the long-term impacts of stresses that might accompany pervasive monitors are cumulative, or if they wear off over time. Future research should investigate these long-term concerns.

Conclusions

The effects of using computerized performance monitors should be studied due to the wide-ranging potential impacts on worker stress and performance. Previous literature has relied on survey data and case studies for understanding the phenomenon. This study made use of an experimental setting, and involved measures of performance as well as measures from physiological and perceptual instruments.

In sum, the physiological measures did not seem to differ when comparing the computer-monitored group with the control group. An exploratory human-monitored group also did not seem to produce any striking results. Nevertheless, pulse rates appeared to vary enough to warrant their use in future studies. In contrast, skin temperature levels appeared too stable for highly precise measurement in such an experimental setting.

Perceived stress levels provided interesting results. In settings where a computerized performance monitor would make sense, there was less state anxiety than where such monitors would not make sense. Similar findings were apparent for the use of a human monitor.

Work quality appeared to suffer as a result of using a computerized monitor, but the speed of completion increased sharply. That is, efficiency appeared to increase to a striking degree, but
only because the amount of time spent decreased much more quickly than did the overall quality. If the highest quality is necessary, then a monitor might not be desirable.

Future research should make use of larger sample sizes, more types of tasks, additional and automated physiological measures, and focus on a longer-term perspective. Such research is necessary to investigate more fully some of the important problems solved, as well as caused by computerized performance monitors.
VII. REFERENCES


