Effects of Supervisory Monitoring on Productivity and Quality of Performance

Neil Brewer and Tim Ridgway
The Flinders University of South Australia

Effects of supervisory monitoring on performance quantity and quality were examined. Experiment 1 participants performed 2 tasks and were monitored periodically on 1 or both tasks, with outputs either identifiable or unidentifiable. Experiment 2 compared several monitoring procedures varying in quantity emphasis. In Experiment 1, an apparent quantity focus positively affected quantity but created quantity and quality decrements on the unmonitored task. Experiment 2 indicated that these effects were most pronounced when monitoring explicitly emphasized quantity. Despite the positive effect on the monitored task, monitoring had a negative influence on a composite quantity-quality performance measure, with this interpreted in terms of changes in participants’ speed-accuracy criteria. Some important implications for supervisory practices in organizational settings are discussed.

Performance monitoring is a term used to describe a range of behaviors that supervisors in organizational contexts use to keep track of how their subordinates are performing. Monitoring can encompass supervisors observing employees at work, examining their work products or outputs, asking them to provide verbal reports on their progress, checking personnel files or records, or even soliciting reports from another employee or supervisor. Many models of the leadership process (e.g., Fleishman, 1953; Komaki, Zlotnick, & Jensen, 1986; Luthans & Lockwood, 1984; Yukl, 1994) identify performance monitoring as a key element of the supervisory process, and recent studies of the behavioral characteristics of effective supervisors (Brewer, Wilson, & Beck, 1994; Komaki, 1986; Komaki, Desselles, & Bowman, 1989) have shown that supervisory effectiveness and the time spent monitoring are positively correlated.

The significance that is attributed to monitoring in the supervision process is derived from the fact that monitoring provides a mechanism whereby performance consequences can be provided contingent on employee performance. Monitoring employees’ performance provides supervisors with the information or data that are the basis for positive recognition and corrective feedback and for the administration of organizational rewards and punishments. Given the well-established relationships between contingent performance consequences and behavior, effective monitoring seems likely to play an important role in the development and maintenance of effective work performance.

It is not uncommon in the real world, however, for monitoring to occur without any accompany-
ing consequences. Further, the argument has recently been advanced that monitoring—even in the absence of any associated consequences—may significantly affect work performance. In developing this argument, parallels have been drawn between monitoring and the extensively researched social facilitation phenomenon. Supervisory monitoring in a social organizational context typically involves employees working for extended periods of time with at least some brief periods spent in the presence of the supervisor. The social facilitation paradigm typically involves contrasting an individual’s performance in the presence of another (audience or coactor) with their performance while acting alone (Guerin, 1993). There are some obvious differences between the two situations. For example, supervisory monitoring is usually enacted in a context where work periods are much longer than the very brief periods (e.g., commonly a few minutes and seldom more than 20–30 min) that characterize social facilitation experiments. More often, monitoring also involves periodic rather than continued supervisory presence, and the supervisor is likely to be perceived as having power over the administration of organizational rewards and punishments that the coactor in the social facilitation paradigm does not. Nevertheless, the parallels are such that theorizing and research on the social facilitation paradigm has provided a useful starting point for thinking about the likely effects of supervisory monitoring.

Extensive research within the social facilitation paradigm has shown that (a) on simple or well-learned tasks where the dominant response is typically the correct one, performance improves in the presence of others and (b) on complex or unlearned tasks where the dominant response is likely to be incorrect, performance deteriorates in the presence of others (for a comprehensive review, see Guerin, 1993). A recent group of studies by Aiello and his colleagues (Aiello & Kolb, 1995; Aiello & Shao, 1993; Aiello & Svec, 1993; Kolb & Aiello, in press) that investigated the effects of electronic or computer monitoring is included among this research. Consistent with the social facilitation literature, these studies have shown that when participants believed that their performance was being monitored electronically, productivity on simple tasks (or of skilled performers) was enhanced, whereas performance on complex tasks (or of unskilled performers) was impaired. An important characteristic of all these studies was that (as for much of the social facilitation literature) work sessions were generally very brief and supervisory monitoring was continuous. In another recent study (Larson & Callahan, 1990) involving longer work sessions and periodic supervisory monitoring (checks on output every 20 min), participants working at two different tasks, and monitored on one, produced more on the monitored task than on the unmonitored task. However, contrary to what might have been expected from the social facilitation perspective, no overall effects on productivity were noted when monitored groups were compared with nonmonitored controls.

There are a number of important issues—methodological, theoretical, and practical—that arise from these recent studies of the effects of performance monitoring. First, from a methodological perspective, analyses in some recent studies have focused on the effects of monitoring on indexes of productivity such as total output or number of correct responses (e.g., Aiello & Kolb, 1995). In others, the dependent variables of performance quantity and performance quality have been analyzed separately (e.g., Aiello & Svec, 1993; Larson & Callahan, 1990), the latter strategy sometimes justified (e.g., Larson & Callahan, 1990) on the basis of a negligible correlation between the two dependent variables. Both approaches may result in a failure to detect important monitoring-induced differences in the regulation of speed and accuracy. Effects on quantity (speed) and quality (accuracy) depend on how different individuals (within, as well as between, conditions) trade off these two parameters. Research on the speed–accuracy trade-off has established that individual participants who perform under instructions that emphasize both speed and accuracy (or quantity and quality) can—depending on the position on the speed–accuracy trade-off function they adopt—produce large changes in accuracy with only very minor variations in responding speed (e.g., Brewer & Smith, 1984, 1989; Pachella, 1974). This research has also shown that participants can vary their speed of responding substantially with minimal (or no) effects on accuracy. Most of the analysis strategies typically used by researchers
confronted with dependent measures of both speed and accuracy (e.g., ignoring error responses and using analysis of covariance to adjust speed data for accuracy) are problematic (see Pachella, 1974). One recommended solution to the problem of analyzing speed and accuracy data involves treating speed and accuracy as multivariate-dependent variables (Pachella, 1974). The implication of this work on speed-accuracy regulation for the present context is that a more complete understanding of the effects of different monitoring conditions on performance may only be provided by examining their impact on individuals' productivity and quality as a composite or multivariate-dependent variable. This strategy was used in the present study.

Second, from a theoretical perspective, it is important to identify the mechanism(s) underlying any effects of monitoring. Theorizing on the social facilitation phenomenon provides many possible starting points. A number of theoretical accounts of this phenomenon have been advanced; the mere presence, evaluation apprehension, objective self-awareness, self-presentation, and distraction–conflict accounts (for overviews of these and other positions, see Guerin, 1993) are some of the better known. Perhaps the most obvious place to start is suggested by theorizing about the evaluation concerns or apprehension induced by the presence of another party. For example, Cottrell (1972) argued that the presence of other people in a social situation may lead to positive or negative evaluations of the way people behave. Through experience, people learn to anticipate or expect such evaluations, and this anticipation (concern or apprehension) has an arousing effect on performance. It has been argued that this increased arousal produces social facilitation effects by increasing the likelihood of dominant-response emissions (i.e., correct responses on simple or well-learned tasks and incorrect responses on complex or unlearned tasks). It has been difficult to unequivocally establish that evaluation concern or apprehension mediates the social facilitation phenomenon, principally because of the difficulty in establishing a no-evaluation experimental control condition that still involves having another person present during performance (see Guerin, 1993). Nevertheless, it is clear from the social facilitation literature that effects on performance are likely to be found if participants perceive that the other party (e.g., an experimenter) is evaluating their performance (Guerin, 1993, p. 145).

The parallels between employees in the organizational context and participants in the social facilitation experiment are that both groups would typically expect their completed work products to be evaluated by the relevant supervisor or experimenter. Thus, in both cases some level of ongoing evaluation concern is to be expected. However, whether or not supervisory-monitoring effects in an organizational context parallel those of the ever-present evaluative experimenter during the brief social facilitation experiment is of particular interest. It might be expected that supervisory monitoring would signal an increased evaluative focus on the part of the supervisor and this would translate into increased evaluation concerns for the employees. In turn, these increased evaluation concerns should be reflected in enhanced performance on relatively simple or well-learned tasks or in impaired performance on complex or unlearned activities.

Two previous studies have suggested that increased evaluation concern may contribute to the effects of supervisory monitoring on performance. As previously noted, Larson and Callahan (1990) found a clear difference between productivity on monitored and unmonitored tasks when participants juggled two different tasks, only one of which was monitored. This difference disappeared, however, when they controlled for participants' perceptions of the importance of doing well on the respective tasks. Perceived importance of doing well on a task is one way of operationalizing evaluation concern in a self-report measure. In another recent study using a paradigm similar to Larson and Callahan's, Brewer (1995) found that participants who were working in a group setting and being monitored individually (on one of two tasks that they were performing) perceived that the supervisor was evaluating and comparing their performance, and therefore, increased their productivity on the monitored task at the expense of the unmonitored task. In contrast, when the outputs of the group as a whole were monitored by the supervisor, participants did not report the same closeness of evaluation or make the same performance adjustments. This study also found that perceived evaluation concern mediated the adjustments in productivity.
between monitored and unmonitored tasks. However, the absence of an unmonitored control condition in that study prevented an assessment of monitoring's impact (and mediating effects of evaluation concern) on overall performance (i.e., quantity and quality across both monitored and unmonitored tasks).

We conducted two experiments that examine this issue further. In Experiment 1, we examined monitoring effects on a composite speed-accuracy variable under conditions intended to maximize or minimize any evaluation concerns that monitoring may have induced. In Experiment 2, we examined monitoring effects on performance-quantity and performance-quality outputs under different monitoring conditions that varied in the extent to which output quantity was emphasized.

Experiment 1

Experiment 1 examined whether the level of evaluation concern that participants perceived mediated effects of monitoring on performance. The participants were "employed" under the supervisory control of an experimenter and worked on what we considered to be relatively well-learned tasks under experimental manipulations designed to either maximize or minimize any evaluation concern that monitoring may have induced. Unlike most studies conducted within the social facilitation framework where (a) the experimental sessions often last only a few minutes and seldom more than 20-30 min and (b) the experimenter or coactor is present throughout, the experimental session in Experiment 1 lasted 2 hr and the supervisor or experimenter only monitored performance periodically. Although this is still a long way from the parameters of the real-world supervisory context, it represents an important extension of earlier studies in which work periods have been brief.

Using the paradigm that Larson and Callahan (1990) first outlined, we examined precisely how periodic monitoring affected the quantity and quality of subordinate performance. Participants were given two different tasks during the 2-hr session and were instructed to work as rapidly and accurately as possible on both tasks. They worked under one of four monitoring conditions: control condition (they worked uninterrupted for the session), periodic supervisory monitoring on one task, periodic monitoring on the other task, or periodic monitoring of both tasks. We crossed this experimental manipulation with a manipulation of task-output identifiability that was designed to enhance the manipulation of evaluation concern. Participants worked under conditions where their task output was either (clearly) identifiable as their own by the supervisor or (apparently) unidentifiable. For half of the participants in each monitoring condition, the supervisor could clearly identify the source of the work output on each task; for the others, the source of the work output was apparently unidentifiable.

We hypothesized that when only one task was monitored periodically, monitoring would increase evaluation concern and, in turn, performance (quantity, quality, or both) on the monitored task relative not only to the unmonitored task but also to the unmonitored controls. When both tasks were monitored, we expected increased evaluation concern and performance (quantity, quality, or both), relative to the unmonitored controls, on both tasks. However, it was also expected that these effects would be more pronounced when participants believed that their work outputs were clearly identifiable and less pronounced when participants believed that their outputs were unidentifiable because, under the former conditions, increased emphasis on evaluation would likely be translated into evaluation concerns. Finally, it was expected that participants' perceived levels of evaluation concern would mediate the effects of monitoring on performance.

Method

Participants

Participants were 192 undergraduate students (80 male and 112 female) who were recruited through the student employment service and paid $30 for their participation. Their ages ranged from 17-50 years ($M = 23.1$). Participants were randomly assigned to eight experimental groups (24 per group); there were no significant differences between groups in either age or gender composition. The supervisor was a 30-year-old man who was unaware of the aims of the experiment and was trained to deliver instructions and to monitor in a standardized manner.
**Procedure**

**Pretest.** Participants completed pretests for reading speed and spelling. Reading speed was measured with the reading subtest of the Cooperative Reading Comprehension Test—Form Y (Australian Council for Educational Research, 1973). The spelling pretest was the identical format of the one Larson and Callahan (1990) used. Participants were advised at the pretest that they were participating in an investigation of factors affecting the efficiency with which people perform clerical tasks.

**Experimental tasks.** The tasks performed in the experimental session also matched the format of those Larson and Callahan (1990) used. One task involved proofreading a lengthy psychology dissertation and listing spelling or typographical errors that had been inserted into the text. In the other task (an alphabetizing task), participants were provided with many sheets of paper, each containing a list of 15 common words. On each sheet, participants had to sort the words into two columns on the basis of each word's first letter and then sort the two columns into a single list in alphabetical order. We gave all participants specific instructions on the two tasks to ensure that they understood the requirements.

**Experimental design and manipulations.** The study used a 4 (supervisory behavior) X 2 (output identifiability) X 2 (task) design, with task as a within-subjects factor. Each participant worked at two tasks under one of four different supervisory-behavior (i.e., monitoring) conditions, with either identifiable or unidentifiable task output. Participants attended experimental sessions with 5-6 other participants, but each participant worked alone in a separate cubicle. Within any experimental session, we exposed all participants to the same manipulation.

We gave all participants general instructions indicating that for the next 2 hr, they would be working on two tasks, that they could switch back and forth between the two tasks as they wished, that the goal was to work as fast and as accurately as possible on both tasks, and that the supervisor would be evaluating how well they did on both tasks at the end of the session.

Participants worked under one of four levels of the supervisory-behavior variable. In each of these, we designed the initial instructions to give participants the impression that at the end of the session, the supervisor would evaluate their performance on the two tasks. One level involved nothing more than the participants working uninterrupted for the entire session. In the other three levels, the supervisor periodically entered the cubicles and checked on the participants' progress. Upon entering the cubicle (at approximately 20-min intervals, on five occasions during the session), the supervisor said (depending on which task was designated as the monitored task), “Let's see how you are going on the alphabetizing/proofreading task,” picked up the completed output from the appropriate task, leafed through it for about 45 s, and left the room. For one level, proofreading was the monitored task; for another, alphabetizing was monitored; and for the remaining level, both tasks were monitored. When both tasks were monitored, we checked outputs for both tasks at each monitor and we varied the order in which they were checked randomly from one monitor to the next. If participants made comments to the supervisor during a monitor (e.g., reporting how much they had done), the supervisor made no response other than “hm,” “aha,” or a nod of the head.

Participants also worked under one of two levels of output identifiability. In the identifiable condition, all task-record sheets in each participant's cubicle were personally labeled and a personally labeled envelope was provided for participants' completed record sheets. In the unidentifiable condition, task-record sheets were unlabeled and we instructed participants not to enter their names on them. An unlabeled envelope was provided for their completed record sheets and we showed participants a large box that all participants who attended that session were to place their unlabeled envelopes in at the end of the session. We coded the data records (by slight variations in the lengths of line rules on each page) so that the supervisor could link each page to the particular cubicle the participant had worked in and, in turn, to the participant.

After participants completed the 2-hr work session, we gave them a questionnaire that was designed to assess their attitudes toward the supervisor, their awareness of the supervisor's monitoring, the extent to which they perceived that the supervisor could personally evaluate them, and the extent of their evaluation concern.
We tapped monitoring awareness through participants’ estimates of the frequency with which the supervisor checked on their progress in each task. Perceptions of close, personal evaluation likelihood were reflected in participants’ responses to four items (two equivalent items for each task) that assessed the extent to which they perceived the supervisor could accurately assess their personal output on each task and the extent to which they perceived the supervisor could reliably compare their personal performance on each task with that of other participants. Responses to each of these items were made on an 11-point scale (1 = not at all; 11 = very much). Participants indicated evaluation concern on a 7-point scale (1 = not at all important; 7 = very important) that tapped the extent to which they perceived that it was important to perform well on each task.

Dependent measures. We obtained performance-quantity and -quality measures for both tasks. Alphabetizing quantity was the total number of words alphabetized; alphabetizing quality was the percentage of words placed in correct order. Proofreading quantity was the total number of lines read; proofreading quality was the number of misspelled words identified — any correct words (inappropriately) identified/total lines read. Because comparison of raw scores for the two different tasks was not meaningful, we converted scores for both tasks to standard scores on the basis of the means and standard deviations for all participants on each measure.

Results and Discussion

Manipulation Checks

Participants’ reactions to the supervisor and perceptions of the monitoring and output-identifiability manipulations were analyzed in separate 4 (supervisory behavior) × 2 (output identifiability) × 2 (task) analyses of variance (ANOVAs), followed (where appropriate) by post hoc contrasts (Tukey a, p < .05). Participants’ feelings about the supervisor were checked by three 7-point bipolar scales (annoying—reassuring, nervewracking—relaxing, and distracting—helpful). Responses were on the positive side of the scale midpoint in all conditions, indicating that the supervisor was perceived in a positive manner across all conditions; there were no significant main or interaction effects on these variables.

In response to questions concerning the frequency with which the supervisor checked on their progress on each task, there was a significant effect for supervisory behavior, F(3, 184) = 66.46, p < .001, and a significant Supervisory Behavior × Task interaction, F(3, 184) = 83.85, p < .001. The unmonitored control group reported less frequent monitoring (M = 0.50, SD = 1.10) than any of the monitored groups (Ms = 2.63–3.42, SD = 0.76–1.43). More important, perceived frequency of monitoring for monitored tasks exceeded that for unmonitored tasks (M = 3.87 vs. 0.99, SD = 1.35 vs. 1.52). Further, participants in the monitoring conditions perceived that the alphabetizing task was monitored more frequently than the proofreading task, regardless of whether one or both tasks were being monitored (M = 4.19 vs. 3.33, SD = 1.32 vs. 1.54).

For many participants, however, the perceived frequency of monitoring did not match the actual frequency (5 checks, in addition to the final evaluation). Consequently, we trimmed the full data set by deleting participants whose perceptions of monitoring frequency suggested that their monitoring-manipulation perceptions may have been grossly inaccurate. Participants were deleted if they reported (in any condition) that an unmonitored task was checked more than twice or a monitored task was checked twice or less. This reduced the sample size from 192 to 142, with cell sizes for the identifiable and unidentifiable conditions, respectively, 22 and 23 (unmonitored), 22 and 19 (alphabetizing monitored), 17 and 15 (proofreading monitored), and 12 and 12 (both tasks monitored). It is not surprising that the contrast between perceived monitoring frequency for monitored and unmonitored tasks was more striking (M = 4.42 vs. 0.38, SD = 1.08 vs. 0.67). It was also the case that participants in the monitoring conditions still perceived that alphabetizing was being monitored more frequently than proofreading, whether one or both tasks were monitored (M = 4.63 vs. 4.18, SDs = 1.15). Thus, for the trimmed sample, perceived monitoring frequency was similar to the actual frequency, although monitoring of proofreading was not as salient as it was for alphabetizing. The reported
SUPERVISORY MONITORING, PRODUCTIVITY, AND QUALITY

Table 1
Perceptions of Personal Evaluation Associated With Alphabetizing and Proofreading Performance Under Different Supervisory-Behavior and Output-Identifiability Conditions in Experiment 1

<table>
<thead>
<tr>
<th>Supervisory behavior</th>
<th>Output and task</th>
<th>Control</th>
<th>Alpha monitored</th>
<th>Proof monitored</th>
<th>Both monitored</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled</td>
<td>Alphabetizing</td>
<td>8.98</td>
<td>9.36</td>
<td>7.44</td>
<td>8.88</td>
<td>8.71</td>
</tr>
<tr>
<td></td>
<td>Proofreading</td>
<td>9.26</td>
<td>8.00</td>
<td>8.79</td>
<td>7.96</td>
<td>8.55</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>9.12</td>
<td>8.68</td>
<td>8.11</td>
<td>8.42</td>
<td>8.63</td>
</tr>
<tr>
<td>Unlabeled</td>
<td>Alphabetizing</td>
<td>5.59</td>
<td>6.08</td>
<td>4.43</td>
<td>5.83</td>
<td>5.51</td>
</tr>
<tr>
<td></td>
<td>Proofreading</td>
<td>5.11</td>
<td>4.29</td>
<td>5.70</td>
<td>5.79</td>
<td>5.13</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>5.35</td>
<td>5.19</td>
<td>5.07</td>
<td>5.81</td>
<td>5.32</td>
</tr>
</tbody>
</table>

Note. Alpha = alphabetizing task; Proof = proofreading task.

analyses were subsequently conducted on the trimmed data set.

We assessed the effectiveness of the output-identifiability manipulation by averaging participants' responses to two items that assessed the extent to which participants perceived that the supervisor could accurately assess their personal output on each task and the extent to which they perceived that the supervisor could reliably compare their personal performance on each task with that of other participants. As expected, there was a significant main effect for output identifiability, $F(1, 133) = 48.91, p < .001$, with participants who worked with labeled output perceiving closer personal evaluation than those who worked with unlabeled output ($M = 8.63$ vs. $5.32$, $SD = 2.61$ vs. $3.10$). But, contrary to expectations, the absence of any significant interactions involving the output-identifiability variable indicated that periodic monitoring did not lead to perceptions of heightened personal evaluation, even when personal outputs were clearly identifiable. There was, however, a significant Supervisory Behavior $\times$ Task interaction, $F(3, 133) = 11.91, p < .001$. As shown in Table 1, participants perceived lesser degrees of personal evaluation relative to unmonitored controls, on the unmonitored task when only one task was monitored. Perceptions of personal evaluation on monitored tasks were not, however, heightened by monitoring.

Table 2
Evaluation Concern Associated With Alphabetizing and Proofreading Performance Under Different Supervisory-Behavior Conditions in Experiment 1

<table>
<thead>
<tr>
<th>Supervisory behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Alphabetizing</td>
</tr>
<tr>
<td>Proofreading</td>
</tr>
</tbody>
</table>

Note. Alpha = alphabetizing task; Proof = proofreading task.

Evaluation-Concern Measure

Items that assessed the extent to which participants perceived that it was important to perform well on each task probed participants' evaluation concerns. A 4 (supervisory behavior) $\times$ 2 (output identifiability) $\times$ 2 (task) ANOVA on this measure revealed only a significant Supervisory Behavior $\times$ Task interaction effect, $F(3, 133) = 6.20, p = .001$. As shown in Table 2, evaluation concern associated with the monitored task increased in relation to concerns reported by unmonitored controls, when only one task was monitored. When both tasks were monitored, evaluation concern associated with the alphabetizing task was higher than in the unmonitored
control condition; for the proofreading task, this difference did not reach significance.

**Performance Measures**

Separate 4 (supervisory behavior) \(\times\) 2 (output identifiability) ANOVAs were conducted on the reading-speed and spelling pretest measures. Although there were no differences between groups on the spelling measure, a significant effect for supervisory behavior: \(F(3, 134) = 4.60, p < .01\), indicated differences between groups on the reading-speed measure. In the subsequent multivariate analysis of variance (MANOVA) on performance data, reading speed was entered as a covariate.

A 4 (supervisory behavior) \(\times\) 2 (output identifiability) \(\times\) 2 (task) MANOVA was conducted on the dependent variables of standardized performance quantity and quality. Table 3 shows the summary statistics for the quantity and quality scores for the various conditions. (Participants maintained overall accuracy levels of 99% and 64% for alphabetizing and proofreading, respectively.) Neither the supervisory-behavior nor the output-identifiability main effects were significant, nor was the interaction between these variables. There was, however, a significant Supervisory Behavior \(\times\) Task interaction, \(F(3, 134) = 6.58, p < .001\). Inspection of the data in Table 3 suggests some parallels between the performance-quantity and the evaluation-concern data: Namely, increased productivity on the monitored task when only one task was monitored, increased productivity on the alphabetizing task when both tasks were monitored, and increased quality of proofreading performance when that was the monitored task. However, there were also some notable differences. Productivity on the unmonitored task when one task was monitored showed a decline relative to unmonitored controls, whereas evaluation concern did not. Proofreading productivity when both tasks were monitored also declined, whereas reported evaluation concern in this condition increased. And when both tasks were monitored, there was a suggestion that performance quality on both tasks decreased, despite increased evaluation concerns. (Separate univariate ANOVAs and post hoc testing on performance quantity and quality, with reading speed as a covariate, confirmed the statistical significance of the above trends in the productivity data; the trends described in the performance-quality data, however, were not statistically significant.)

Given the previously outlined pitfalls associated with the independent analysis of speed (quantity) and accuracy (quality) data, we further examined the effects of the experimental manipulations on a composite speed and accuracy dependent variable and whether or not perceptions of evaluation concern mediated such effects through a latent-variable path analysis. In this context, the performance variable was constructed as a linear

<table>
<thead>
<tr>
<th>Supervisory behavior</th>
<th>Control</th>
<th>Alpha monitored</th>
<th>Proof monitored</th>
<th>Both monitored</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>P</td>
<td>A</td>
<td>P</td>
</tr>
<tr>
<td><strong>Quantity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labeled</td>
<td>-0.50</td>
<td>.02</td>
<td>0.13</td>
<td>-.07</td>
</tr>
<tr>
<td>Unlabeled</td>
<td>-0.09</td>
<td>-.06</td>
<td>0.96</td>
<td>-.61</td>
</tr>
<tr>
<td>Overall</td>
<td>-0.29</td>
<td>-.02</td>
<td>0.52</td>
<td>-.32</td>
</tr>
<tr>
<td><strong>Quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labeled</td>
<td>-0.04</td>
<td>-.18</td>
<td>0.21</td>
<td>-.06</td>
</tr>
<tr>
<td>Unlabeled</td>
<td>0.08</td>
<td>.12</td>
<td>0.04</td>
<td>.04</td>
</tr>
<tr>
<td>Overall</td>
<td>0.02</td>
<td>-.03</td>
<td>0.14</td>
<td>-.01</td>
</tr>
</tbody>
</table>

*Note. Alpha = alphabetizing task; Proof = proofreading task; A = alphabetizing; P = proofreading.*
more than two jackknife standard errors (values in parentheses). As shown in Figure 1, meaningful direct paths were formed from both pretest and supervisory behavior to performance, with these explaining 5.0% and 12.4% of the variance in performance, respectively. Supervisory behavior also exerted an influence on performance through the perceived evaluation-concern variable, although this effect was much weaker than the direct effect. Perceived evaluation concern explained 11.9% of the variance in performance, but the supervisory-behavior variable explained only 10.3% of the variance in evaluation concern.

However, for both the direct and indirect effects, the net effect of supervisory behavior on performance was negative in nature. Examination of the manifest–latent variable relationships in conjunction with the inner model coefficients illustrates the specific nature of the monitoring effects. As Figure 1 indicates, monitoring proofreading had a direct positive impact on overall
performance (i.e., negative weight on supervisory behavior followed by a negative path to performance) and also a lesser indirect effect via increased evaluation concern for proofreading performance (negative path to evaluation concern with a positive loading on evaluation concern from the proofreading task) relative to evaluation concern for alphabetizing. Enhanced productivity and quality of performance on the proofreading task reflected these effects, though not without a decrement in alphabetizing productivity. However, these gains were outweighed by the effects on overall performance when either alphabetizing or both tasks were monitored. In both cases, direct and indirect (through evaluation concern) negative effects on performance were detected. Although alphabetizing quantity increased (negative path to performance followed by negative loading on alphabetizing quantity), this was at the expense of both productivity and quality of performance on the proofreading task. Thus, compared with unmonitored controls, the net effect of the various monitoring manipulations was a reduction in performance.

The general trend in the evaluation-concern data was increased concern on the monitored task (relative to unmonitored controls), regardless of the monitoring condition. There was also an effect of supervisory behavior on performance that evaluation concern mediated. As predicted, evaluation concern was heightened on whatever task was perceived as being monitored more frequently within each condition, and this was translated into enhanced performance on that particular task (at the expense of the other). However, the effect was weak and less influential than the direct effect of the supervisory-behavior variable on performance.

It is the nature of this latter, direct effect that is particularly interesting and in need of explanation. In particular, why should the impact of monitoring on performance differ depending on whether it focused on proofreading, alphabetizing, or both tasks? A possible post hoc explanation relates to differences in the way that participants may have perceived the focus of monitoring under the different conditions. Monitoring involved the supervisor leafing through completed output for about 45 s. When proofreading was being monitored, participants may have interpreted this as indicative of interest in proofreading quantity, quality, or both (because participants may have perceived this kind of output inspection as providing the supervisor with a guide as to the number of pages completed and at least some guide as to the accuracy of any corrections that had been made, although probably not the proportion of mistakes actually detected). Monitoring alphabetizing in this same way, however, would convey the impression that the supervisor was predominantly interested in quantity (indicated simply by the number of pages scanned) because without the supervisor taking the time to closely inspect the various columns on each page, participants could reasonably assume that the supervisor was gaining little appreciation of quality as he or she leafed through the output. The pattern that emerged when both tasks were monitored was similar to the pattern that was obtained when alphabetizing was monitored, and this pattern most likely reflects the fact that for some reason, participants perceived that supervisors monitored alphabetizing significantly more frequently than proofreading, even when the supervisor monitored both tasks at each visit.

Two obvious questions arose from these suggestions. First, could systematically manipulating the focus of monitoring so as to contrast the effects of monitoring that explicitly targeted only quantity of performance with monitoring that did not clearly favor quantity over quality reproduce effects such as those in Experiment 1? Second, if we could reproduce such effects, could we suggest some plausible explanation for the fact that the overall or net effect on performance was negative (rather than positive) in nature?

Experiment 2

In Experiment 2, we attempted to reproduce the patterns of performance detected under monitoring in Experiment 1. The aim was to test the possibility that the different patterns of monitoring effects we observed across tasks in Experiment 1 reflected a differential emphasis on performance quantity. We contrasted the effects of a number of different methods of monitoring, with these methods varying in terms of the extent to which performance quantity (or productivity) was emphasized.

In operationalizing methods of monitoring, we were guided by previous research focusing on
how supervisory monitoring is typically enacted in field settings. For example, behavioral studies of the monitoring patterns of field supervisors (e.g., Brewer et al., 1994; Komaki, 1986; Komaki et al., 1986) have distinguished between procedures such as work sampling (where supervisors actually observe the employee performing a task), product sampling (checking on work products or output), and soliciting self-reports (asking employees about their performance). Here we contrasted four such procedures with an unmonitored control condition. Two monitoring conditions focused explicitly on productivity or output quantity and involved either product sampling or making a quantity-related inquiry of subordinates (the self-report: specific condition). The other two did not explicitly distinguish between quantity and quality and involved either work sampling or making a general inquiry of subordinates (the self-report: general condition). We examined how these different types of monitoring affected the quantity and quality of subordinate performance using a similar paradigm to that used in Experiment 1.

Method

Participants

Another 150 paid undergraduate students (71 male and 79 female) from the employment service were assigned to five groups (30 per group). Participants' ages ranged from 18–44 years ($M = 23.2$), and experimental groups did not differ significantly in age or gender composition.

Procedure

Experimental design and manipulations. The pretests, experimental tasks, general experimental protocols, and dependent measures were the same as in Experiment 1. The design was a 5 (monitoring type) $\times$ 2 (presence of monitoring) $\times$ 2 (task monitored) design, with presence of monitoring as a within-subjects factor. Each participant worked at two tasks under one of five different monitoring conditions (including an unmonitored control condition), with one task monitored periodically by the supervisor and the other not monitored. We gave participants the same general instructions as in Experiment 1, with these instructions emphasizing speed and accuracy on both tasks, and we told participants that the supervisor would be evaluating how well they did on both tasks at the end of the session.

Control-group participants worked undisturbed for the whole session. Participants in the four monitored groups were individually checked by the supervisor. For half of the participants in each group, proofreading was the monitored task; for the other half, alphabetizing was the monitored task. The different methods of monitoring were enacted as outlined below.

Work sampling. After 20 min, the supervisor entered the cubicule, sat down next to the participant, and watched the participant working on the task they were performing at that time for 1 min. This task was then designated as the monitored task for that participant. Subsequently, after intervals of about 15 min, the supervisor checked (unobtrusively) from the corridor outside the cubicles to determine what task participants were working on. If the participant was working on the monitored task, the supervisor would again enter the room (as close as possible to the 20-min markers), sit down, and watch the participant work for 1 min. If the participant was working on the other task, the supervisor checked again from the corridor every couple of minutes and only monitored the work after the participant had switched to the monitored task. Although the target in this condition was to monitor performance on one task every 20 min, the actual pattern varied around this target from participant to participant. After the first monitor at the 20-min mark, performance was monitored, on average, after 48 (range = 38–67), 71 (52–97), 93 (75–110), and 110 (95–117) min.

Product sampling. The supervisor entered the room at 20-min intervals and collected the output from the monitored task. When collecting the output, the supervisor said, “Let’s see how many pages you have proofread” or “Let’s see how many lists you have sorted.”

Self-report: general. The supervisor entered the room at 20-min intervals and (depending on which task was designated as the monitored task) inquired, “How are you progressing with the proofreading/alphabetizing task?”

Self-report: specific. The supervisor entered the room at 20-min intervals and (depending on
which task was designated as the monitored task) inquired, “How are you progressing with the proofreading task? How many pages have you proofread?” or “How are you progressing with the alphabetizing task? How many lists have you sorted?”

After completing the 2-hr work session, participants were given a questionnaire that explored their awareness of the supervisor’s monitoring, their reaction to the supervision experienced, and the extent of their evaluation concern, again indicated by the extent to which they perceived that it was important to perform well on each task.

Results and Discussion

Manipulation Checks

Participants’ feelings about the supervisor were again checked as in Experiment 1, with responses on the positive side of the scale midpoint and no significant differences between groups. Participants in all monitored groups reported significantly more frequent monitoring ($M_s = 4.9-5.4$, $SD_s = 1.11-1.25$) than did control participants ($M = 1.2$, $SD = 0.94$), with impressions of monitoring frequency close to the actual frequency (i.e., 5). For a few participants, the perceived monitoring frequency again did not match the actual frequency; using the same criterion as in Experiment 1, we deleted 10 participants (3 from both the control and work-sampling conditions and 2 from each of the self-report conditions) and conducted performance data analyses on the trimmed data set.

Evaluation Concern Measure

Again, items that assessed the extent to which participants perceived that it was important to perform well on each task probed participants’ evaluation concerns. A 5 (supervisory behavior) × 2 (type of task monitored) × 2 (presence of monitoring) ANOVA, with presence of monitoring as a within-subjects factor, revealed only a significant Type of Task Monitored × Presence of Monitoring interaction, $F(1, 123) = 12.65$, $p = .001$. As shown in Table 4, evaluation concern on the alphabetizing task was higher under monitoring than in the control condition except when monitoring involved work sampling. However, this trend was not observed on the proofreading task, perhaps due in part to evaluation concern being closer to the ceiling.

Performance Measures

A 5 (supervisory behavior) × 2 (type of task monitored) × 2 (presence of monitoring) MANOVA, with presence of monitoring as a within-subjects factor and reading speed as a covariate, was conducted on the dependent variables of standardized performance quantity and quality. (Participants in the control group were randomly assigned to monitored and unmonitored task conditions for the purposes of analyses, with the constraint that matching on pretest scores was preserved. Overall accuracy levels were 99% and 58% for alphabetizing and proofreading, respectively.) The effects reported below were independent of type of task monitored: The main effect for type of task monitored did not

<table>
<thead>
<tr>
<th>Supervisory behavior</th>
<th>Work sampling</th>
<th>Product sampling</th>
<th>Self-report: general</th>
<th>Self-report: specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitored</td>
<td></td>
<td>A</td>
<td>P</td>
<td>A</td>
</tr>
<tr>
<td>Unmonitored</td>
<td>5.00</td>
<td>5.69</td>
<td>5.50</td>
<td>5.77</td>
</tr>
<tr>
<td>Overall</td>
<td>5.23</td>
<td>5.82</td>
<td>5.54</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Note. A = alphabetizing; P = proofreading.
approach significance ($F < 1$), nor did any of the interactions between type of task monitored and the other variables. Therefore, the data shown in Figure 2 were collapsed across tasks.

Although the presence of monitoring effect did not reach significance, $F(1, 140) = 2.95, p = .08$, there was a significant Supervisory Behavior $\times$ Presence of Monitoring interaction, $F(4, 140) = 4.12, p < .01$, with the effects of monitoring varying across the monitoring procedure. Examination of the separate data patterns for performance quantity and quality (shown in Figure 2) indicates pronounced differences in performance quantity between monitored and unmonitored tasks when monitoring involved either specific self-reports or sampling of work products, but somewhat lesser effects under the other two procedures where there was no explicit quantity emphasis. No similar, clear trends in the quality data were discernible.

As in Experiment 1, we examined the effects of the different monitoring conditions on the composite speed and accuracy variable and whether evaluation concern mediated any effects by using latent-variable path analysis (with supervisory behavior, pretest performance, perceived evaluation concern, and performance as the respective latent variables). The supervisory-behavior variable was again combined into a latent variable by omitting the unmonitored control as a dummy variable.

Meaningful paths were formed from both pretest and supervisory behavior to performance, with these explaining 9.9% and 9.4% of the variance in performance, respectively (see Figure 3). There was no evidence of an indirect effect of supervisory behavior on performance that was mediated by evaluation concern. As in Experiment 1, the effect of supervisory behavior on performance was negative. Relative to the dummy variable (the unmonitored control group), all methods of monitoring (a) increased performance quantity on the monitored task (i.e., a positive weight on supervisory behavior followed by a negative path to performance, but with a negative loading from monitored task quantity), (b) reduced performance quality on the monitored task (i.e., negative path to performance with a positive loading from monitored-task quality), and (c) reduced performance quantity and quality on the unmonitored task. Thus, although monitored-task quantity improved, it did so at the expense of monitored-task quality and, particularly, both quantity and quality on the unmonitored task, with the net result of overall decline in performance. The weights of the various monitoring conditions indicate that the effects were most pronounced when monitoring involved specific self-reports and product sampling, in other words, when monitoring explicitly emphasized performance quantity. The same type of effect was still evident under the other monitoring conditions where quantity was not explicitly emphasized (i.e., work sampling and self-report: general), but the effect was less pronounced.

---

Figure 2. Mean performance quantity and quality for the monitored and unmonitored tasks in the various monitoring conditions. Work Samp = work sampling; Prod Samp = product sampling; Self-rep:Gen = self-report: general; Self-rep:Spec = self-report: specific.
General Discussion

The two experiments provided converging evidence on the effects of periodic supervisory monitoring on productivity and performance quality. For all monitoring conditions in Experiment 1, productivity increased on the task that was perceived to be monitored more frequently. In one condition, monitored-task quality also increased and despite a fall in productivity on the unmonitored task, net performance increased. However, in the other two conditions, the productivity gain on the task perceived as more frequently monitored was outweighed by the decrements in both productivity and quality on the other task. Overall, the monitoring conditions exerted a negative net effect on a composite-quantity-quality-performance variable. We suspected that the extent to which participants believed that the supervisor's monitoring explicitly focused on quantity may have contributed to the different pattern of results across monitoring conditions. In Experiment 2, we tested this by comparing a number of different methods of monitoring that varied in performance-quantity emphasis. Again, we observed increased productivity on the monitored task accompanied by reduced quality on the monitored task, reduced quantity and quality on the unmonitored task, and a negative net effect on the composite-quantity-quality-performance variable. This effect was more pronounced in the two conditions that explicitly emphasized quantity and less so when there was no specific quantity emphasis. In both experiments, the net effects on performance were of similar magnitude, with 9–12% of the variance in performance explained.

Two important features of these data deserve emphasis. First, participants were clearly discriminating in the sense that they adjusted their performance in a manner that was consistent with the supervisor's monitoring focus, a finding in line with field work on the relationship between monitoring and performance (e.g., Brewer et al., 1994; Komaki, 1986; Komaki et al., 1989).
Second, in adjusting their performance in line with the supervisor's apparent monitoring focus, participants appeared to have difficulty balancing performance quantity and performance quality across both tasks in order to maintain overall levels of performance (i.e., across the two tasks) equivalent to those of unmonitored controls.

The findings failed to support predictions stimulated by the social facilitation literature that effects of monitoring would be mediated by monitoring-induced changes in evaluation concern. Only a very weak mediating effect of evaluation concern was detected in Experiment 1, with no effect detected in Experiment 2. Perhaps evaluation concerns were already as high as could be expected in a context where supervisors were most likely perceived as having limited powers to administer rewards and punishments, or, the self-report measure of evaluation concern may have been insufficiently sensitive. The pattern of results was not consistent with the well-documented social facilitation effects of enhanced performance on simple tasks. For both tasks, the correct response was the dominant response, and yet the overall effect on performance was negative.

Nevertheless, participants' performance patterns were consistent with a behavior-analytic conceptualization of the evaluation-concern phenomenon described in the social facilitation literature (cf. Guerin, 1993, 1994). In these terms, monitoring is seen as providing a discriminative stimulus for particular social consequences. Thus, for example, monitoring one task more frequently than another would signal that a subordinate's increased attention to the former task would be more likely to attract favorable (or forestall unfavorable) consequences from the supervisor. Likewise, monitoring that suggested a greater interest in quantity than quality would signal favorable consequences for increasing quantity at the expense of quality and vice versa. It still remains to be proven that monitoring that focuses explicitly on performance quality boosts quality at the expense of quantity, but the present results lead us to predict such an outcome.

In essence, we are suggesting that monitoring—at least when it is perceived to focus on a particular performance dimension—has an effect by acting as a discriminative stimulus that induces the operator to trade off different dimensions of performance. For example, monitoring that focuses on performance quantity on Task A may induce an individual to trade off quantity (speed) on Task A for quality (accuracy) on Task B. Thus, appropriately focused monitoring is an effective way for a supervisor to bring about a desired change in performance (e.g., increased productivity or increased accuracy). These experiments also suggest that, by inducing the trading of speed and accuracy, monitoring might also net some negative effects on performance that will only be detected by precise measurement of quantity and quality for all task activities.

Basic research on the speed-accuracy trade-off provides some clues as to why this may occur. We know from reaction time experiments that people who are asked to respond as rapidly and accurately as possible over a lengthy series of trials (just as participants in these experiments were asked to do) are unable to maintain ideal fast, accurate response bands just safely above overly fast, inaccurate response bands if they are unable to monitor speed and accuracy of responding with precision (Brewer & Smith, 1984, 1989; Smith & Brewer, 1995). For example, when attempting to respond faster, they are prone to overshoot into overly fast, high-error-rate response bands; or, when attempting to respond slower to avoid overly fast, inaccurate response bands, they are again prone to overshoot into safe (accurate) but overly slow response bands. Further, the nature of the speed-accuracy trade-off function is such that for an individual operating very efficiently (i.e., fast, but still accurate most of the time), (a) attempts to produce even small increases in speed (quantity) are likely to come at considerable cost in terms of accuracy (quality) and (b) attempts to produce any discernible improvement in accuracy will require substantial slowing and, even then, it may be difficult to detect any changes in accuracy (cf. Pachella, 1974). In other words, the overall or net effect on performance (i.e., on a composite speed-accuracy measure) of shifting either way along the trade-off function from an ideal fast, accurate operating point is likely to be negative. Now, when people are performing not one but two tasks (as in these experiments), the number of performance dimensions that they must keep track of increases from two (quantity and quality
on Task A) to four (quantity and quality on Tasks A and B). The likely impact of this is that people will find it much more difficult to monitor speed (quantity) and accuracy (quality) on both tasks with the precision that is necessary to maintain a fast, accurate operating point on both. These appear to be the sort of conditions in which overall negative effects on performance would be detected if supervisory monitoring was enacted in a manner (e.g., explicit quantity or speed focus) that induced individuals to trade accuracy for speed or vice versa. An individual might adjust performance characteristics to match (successfully) expectations suggested by their supervisor's monitoring and in so doing, might actually maintain less effective overall performance.

The hypothesis that overall performance decrements induced by monitoring reflect underlying characteristics of speed-accuracy regulation can be tested by recording latency and accuracy of responses while individuals perform a serial task(s) and by gathering the large amounts of data required to conduct the fine-grained analyses involved in examining the efficiency of speed-accuracy regulation (e.g., see Brewer & Smith, 1984, 1989, 1990; Smith & Brewer, 1995). A similar approach can be applied to the examination of monitoring effects when individuals have only one task to perform. Manipulations that vary the monitoring-focus salience on one particular performance dimension (i.e., quantity or quality) can be examined to determine whether or not they induce corresponding attempts to adjust performance speed or performance accuracy at the expense of the other dimension and at the expense of overall performance.

What can we say about the practical significance of these findings? It is clear that different enactments of supervisory monitoring have the potential to significantly alter the balance between productivity and quality of performance and also to change net performance levels. That is, relatively subtle changes in supervisors' monitoring of subordinates' performances are likely to influence not only how subordinates distribute their efforts across the various activities they perform but also their overall productivity and performance quality. It is likely that many supervisors give little thought to the precise way in which they monitor performance or to the possibility that relatively minor variations in monitoring procedures might have implications for performance. To the extent that these findings generalize beyond the laboratory, these findings have important implications for supervisory practices and training, with supervisors needing to be aware of how their monitoring practices can exert both positive and negative influences on both productivity and performance quality. The rapid growth in the use of electronic or computerized monitoring of work performance makes such considerations even more relevant, particularly because survey research has already demonstrated that operators in computer-monitored environments may perceive a disproportionate emphasis on quantity when compared with quality of performance (e.g., Irving, Higgins, & Safayeni, 1986).

Whether or not these effects would be replicated with different tasks or work activities requires further investigation. The exploration of the effect strength beyond the experimental context is also of interest. Here, participants were unaware of how monitoring would proceed, what its focus was, and so on, and the specific target of the monitoring may not have been detected for two or three monitors, by which time a significant proportion of the work session was complete and there was only a limited time for performance adjustments to take effect. Further, given real-world circumstances where the supervisor has enhanced status and power, including control over the administration of organizational rewards and punishments, it is quite possible that the magnitude of effects would be greater.

Should future studies using methodologies that more closely approximate the realities of work settings confirm the patterns of results reported here, there are a number of obvious and important implications for organizational functioning. Further research that manipulates the precise focus of supervisory monitoring and that is designed to clarify the critical mechanisms underlying the effects of monitoring on speed and accuracy of performance should enhance our understanding of how supervisory monitoring is likely to impact the various dimensions of a subordinate's work performance.

References


Received September 11, 1996
Revision received February 9, 1998
Accepted February 13, 1998