Effects of Group Feedback, Goal Setting, and Incentives on Organizational Productivity

Robert D. Pritchard, Steven D. Jones, Philip L. Roth, Karla K. Stuebing, and Steven E. Ekeberg
University of Houston

While industrial and organizational psychologists have been concerned about the national problem of increasing productivity, most research has focused on simple jobs with the individual as the unit of analysis. Most jobs are more complex and, because of the interdependencies in the work, require group-level interventions and more complex measures of performance or productivity. This research presents a new method of measuring productivity, the Productivity Measurement and Enhancement System (ProMES), and uses measures from this approach as a foundation for group-based feedback, goal setting, and incentives. The experimental design consisted of a baseline period of 8 to 9 months, followed by a 5-month period of feedback based on the ProMES. Goal setting was then added to feedback for 5 months. Finally, incentives in the form of time off from work were added to feedback and goal setting for another 5 months. Results indicated that group-level feedback increased productivity an average of 50% over baseline, group goal setting increased productivity 75% over baseline, and group incentives increased productivity 76% over baseline. Control group data showed no or only a slight increase in productivity over the same period, and level of personnel either stayed the same or decreased. In addition, work attitudes such as job satisfaction, turnover intentions, and morale were as good or better following the interventions. Issues pertaining to group-based interventions are discussed.

Enhancing productivity has been seen as important for our quality of life, our economy, and our competitive position in the world marketplace (Alluisi & Meigs, 1983; American Productivity Center, 1981). In addition, individual organizations are continually concerned about increasing their own productivity in order to improve their operational effectiveness. Although we will discuss the construct of productivity at some length later in this article, for the present, productivity can be considered a group-based concept that deals with the effectiveness or efficiency with which an organization or part of an organization uses its resources to meet its objectives.

Although researchers have shared this concern for enhancing productivity, industrial and organizational psychologists have focused most frequently on only one determinant of productivity—individual performance. One has only to examine such topics as feedback, goal setting, incentives, and performance measurement to see that the predominant focus is on individual performance rather than on productivity as either the efficiency or the effectiveness of a group.

This research on individual performance is valuable, but it has limited applicability for questions concerning productivity. Much of the research has been done on fairly simple jobs; jobs where (a) each person does only a few different tasks, (b) individuals in a unit do largely the same work, (c) the individuals in the unit work independently of one another, and (d) there is a measure of performance readily available.

A limitation with using these simple jobs in our research is that many jobs in actual organizations are more complex than this. In most jobs each person does many things, the different individuals in a unit do very different things, the members of the unit work closely with one another in an interdependent fashion, and there are no ready measures of performance.

This limitation has several important implications. First, much of the research in our field has limited applicability when it comes to its impact on the productivity of organizations. It is not done on the types of jobs that are typical of ongoing organizations. Therefore, the generalizability of our findings to the more prevalent, complex jobs is unclear.

A second implication is that the interdependence between members of the work group in the typical job means that mea-
sures of productivity must be group-based measures. From a practical standpoint, it is impossible to separate out the contributions of each individual when the members must work inter-
dependently to do the work. Furthermore, if one takes the point of view that the output of the whole is greater than the sum of the parts, we would not want to try to separate out individual contributions, even if we could (Steiner, 1972).

Unfortunately, there are no ready ways to measure group productivity. The literature in this area is simply not as fully developed as that for individual performance. The productivity measurement literature offers some suggestions, but the problem of measuring productivity has hardly been solved (Muckler, 1982; Tuttle, 1983). Likewise, the literature on group output measures has no definitive answers (Rowe, 1981). Re-
searchers in organizational effectiveness have worked on the organizational measurement problem for years and cannot agree on what organizational effectiveness is, much less how to measure it (J. P. Campbell, 1977; Goodman, 1979; Steers, 1975; Tuttle, 1983). Thus, we need to refine our conceptualizations and develop high-quality measures in this area.

A third implication of the predominance of complex jobs is that we need to develop group-based interventions in addition to the typical individually based interventions that have been used in the past. Undoubtedly, much of the research on individual interventions will generalize to group-based interventions, but this needs to be explored. Also, we need to determine what special characteristics need to be taken into account in developing group interventions. This is not to suggest that such group interventions have not been studied in the past. The classic experiment by Morse and Reimer (1956) was such a study, and there have been many others as well (cf. Alderfer, 1977; Emmert, 1978; Locke, Shaw, Saari, & Latham, 1981; Nadler, 1979; Thierry, 1987; Woodman & Sherwood, 1980). However, there has been a disproportionately small concern for group productivity in either industrial or organizational psychology.

This line of reasoning led to a research program designed to (a) develop and evaluate a method of measuring productivity for complex jobs and (b) develop and evaluate group-level feedback, goal setting, and incentive interventions. The development and evaluation of the productivity measurement system (termed the Productivity Measurement and Evaluation System, or ProMES) is described elsewhere (Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1986, 1987a). In this article we describe the results of the second objective of the program: to develop and evaluate group feedback, goal setting, and incentive sys-
tems.

Feedback

The positive effect of feedback on performance has become one of the most accepted principles in psychology (see Ammons, 1956; Annett, 1969; Bilodeau & Bilodeau, 1961; Guzzo, Jette, & Katzell, 1985; Ilgen, Fisher, & Taylor, 1979; Nadler, 1979; Sassenwraith, 1975; and M. S. Taylor, Fisher, & Ilgen, 1984). A host of studies have demonstrated the beneficial effects of feedback on individual performance in organizations (e.g., Dockstader, Nebecker, & Shumate, 1977; Ilgen et al., 1979; Ivancevich & McMahon, 1982; Pritchard, Bigby, Beiting, Cowderdale, & Morgan, 1981). In a meta-analysis of psychologically based interventions (Guzzo et al., 1985), feedback was found to have an average effect size of .35 standard deviations (d statistic).

Aside from establishing that feedback works, feedback research has principally followed three lines of inquiry. The first involves investigating the different dimensions of feedback or investigating how a feedback system should be designed along those dimensions so that it can have optimal effect (Beeson, 1973; Berman, Fraser, & Theious, 1970; Chobbar & Wallin, 1984; Christian, 1972; Dockstader et al., 1977; Fisher, 1979; Grelter, 1975, 1980; Grelter & Herold, 1975, 1977; Griswold, 1970; Ivancevich, Donnelly, & Lyon, 1970; Lair & Smith, 1970; Longstreth, 1970; Merrill, 1970; Payne & Hauty, 1955; Pritchard & Montagno, 1978; Pritchard, Montagno, & Moore, 1978; Robinson & Kulp, 1970; Sassenwraith & Younge, 1969; Simpson, 1972; Weidner, 1968; Wike, 1970).

The second line of research has been to examine the effects of moderator variables on these dimensions (e.g., Baron, Cowan, & Ganz, 1974; Baron, Cowan, Ganz, & McDonald, 1974; Baron & Ganz, 1972; Berkowitz, Levy, & Harvey, 1957; Ilgen et al., 1979).


Although most of the feedback research has focused on the impact of feedback on individual performance, some research has been done showing positive effects for groups (e.g., Becker, 1978; Chobbar & Wallin, 1984; Emmert, 1978; Nadler, Cammann, & Mirvis, 1980). Nadler (1979) reviewed the literature on group feedback and presented a model for the effects of group feedback on performance. What literature is available supports the positive effects of group feedback on performance.

Goal Setting

The benefits of combining goal setting with feedback to enhance performance or productivity have been well established in a wide variety of situations (see reviews by Latham & Yukl, 1975b; Locke, 1968; Locke et al., 1981; Steers & Porter, 1974; and Tubbs, 1986). Goal setting has been effectively used with a variety of workers, such as production workers (Dachler & Mobley, 1973), telephone operators (Burke & Wilcox, 1969), logging crews (Latham & Yukl, 1975a), Navy industrial workers (Crawford, White, & Magnusson, 1983), clerical workers and key punch operators (Dockstader et al., 1977; Pritchard et al., 1981), truck drivers (Latham & Baltes, 1975), and engineers (Ivancevich & McMahon, 1982). In their meta-analysis, Guzzo et al. (1985) found goal setting to increase performance with an average effect size of .75 standard deviations (d statistic).

Goal-setting research has followed the same avenues that have been explored in the research on feedback. That is, in addition to work on the effectiveness of goal setting, research has focused on the dimensions of the phenomena, such as goal difficulty, goal specificity, participation in the goal-setting pro-
cess (e.g., Erez & Zidon, 1984; Latham, Mitchell, & Dossett,
1978; Locke et al., 1981; Tubbs, 1986), and the process or mechanisms by which goal setting affects behavior (e.g., Locke, 1968; Locke et al., 1981; Naylor & Ilgen, 1984). There has also been some research on group goal setting (e.g., Becker, 1978; Buller & Bell, 1986; Koch, 1979; Latham & Yukl, 1975b). The general finding is that group goal setting also increases performance.

Incentives

The use of incentives to alter motivation and behavior has been researched for more than half a century (Hull, 1943; Lewin, 1938; Skinner, 1938; F. W. Taylor, 1947; Thorndike, 1911; Tolman, 1932). Incentives have been used in almost every conceivable situation where a behavior is to be influenced. Reviews of incentive motivation theories may be found in J. P. Campbell and Pritchard (1976), DeLeo (1972), Guzzo (1979), and Lawler (1971).

It is quite clear from this literature that financial and nonfinancial incentives can indeed increase performance when the incentive system is properly designed. Guzzo et al. (1985) reported an average effect size for financial incentives of .57 standard deviations with a broad confidence interval that also included zero (−.10 to 1.24). They concluded that the strength of incentive effects depends heavily on the circumstances and methods of applying them.


As with feedback and goal setting, the literature indicates that group-based incentive systems can also be effective (Bullock & Lawler, 1984; Geare, 1976; O'Dell, 1981, 1986; Thierry, 1987.)

It is clear from this literature that feedback, goal setting, and incentives can have positive effects on individual performance. We have also learned a good deal about how they work to influence individual performance. However, much of this large body of research has been done on simple jobs or tasks. As we argued before, many jobs in organizations are not like these simple jobs or tasks.

This implies that we need to work at the group level. We need to develop group-based measures and develop and evaluate group-level interventions (Balk, 1975; Rowe, 1981). What research has been done indicates that these interventions are effective at the group level. However, we do not know nearly as much about the optimal design of such interventions. This research is meant to be a step toward that objective.

Organizational Productivity

Although much has been written on the subject of organizational productivity, there is little consensus concerning its definition (Muckler, 1982). Such a lack of consensus is perhaps not surprising in as much as there are many approaches to and perspectives on productivity. There are, however, several major issues that cut across this literature, and these will be addressed here.

The first issue concerns the overall conceptual approach to be taken. That is, different approaches to measuring productivity are determined by the perspective of those doing the measuring. Tuttle (1983) suggested five perspectives: those of the economist, engineer, accountant, manager, and industrial and organizational psychologist. These approaches differ primarily in what they are trying to learn from the productivity measurement. To the economist, for example, productivity is typically output divided by associated inputs, for example, labor, capital, intermediate products purchased, and time. This approach is typically applied to very large units, such as whole industries or countries, to measure the economic health of those units. This approach includes factors like market conditions, the national economy, and international monetary policies as legitimate factors affecting productivity. In contrast, the industrial-and-organizational-psychologist approach focuses on the personnel subsystem of the organization, and its measures deal with the efficiency or effectiveness of personnel. These approaches are quite different. They measure different things, and they are used for different purposes. They would also result in very different productivity measurement systems.

The second issue is the level of analysis used. In popular literature, one sees the term productivity used to refer to a unit of analysis ranging from the individual to entire countries. In the organizational productivity literature, however, the smallest unit of analysis is typically defined as the work group (Norman & Bahri, 1972; Tuttle, 1983; Tuttle, Wilkinson, & Matthews, 1985).

A third issue revolves around which type of organizational model to use. This issue comes from the organizational effectiveness literature but also applies to productivity. J. P. Campbell (1977) described the two major alternatives as the goal-oriented model and the natural systems model. The goal-oriented model (e.g., Barnard, 1938; Etzioni, 1964; Perrow, 1970) assumes that the organization is run by a set of rational decision makers who have a manageable set of goals for the organization that can be defined well enough to be understood. Organizational effectiveness can be thought of as the degree to which these goals are met. The natural systems model (e.g., Georgopoulos & Tannenbaum, 1957; Katz & Kahn, 1966; Yuchtman & Seashore, 1967) assumes that the demands on an organization are so complex and changeable that it is not possible to identify a finite set of organizational goals that are really meaningful. Instead, the natural systems approach assumes that the overall goal of the organization is survival. To maximize survival, different natural systems models propose or assume that specific characteristics must be maximized. Examples of such characteristics include openness of communication, participation in decision making, and level of organizational trust. Thus, organizational effectiveness is thought of as the degree to which the organization is high on these critical variables.

A third possible organizational model is the multiple constituency approach (e.g., Connolly, Conlon, & Deutsch, 1980; Keeley, 1978; Penning & Goodman, 1977). In this view, the organization is influenced by groups of individuals (constituen-
cies) internal and external to the organization, such as different groups of managers, employees, customers, and so forth. These constituencies have different goals that are based on their own self interests. This model of organizations implies that there is no single set of goals or objectives for the organization.


A fourth issue in the productivity area concerns whether an efficiency or an effectiveness approach should be used in measuring productivity. Both have been proposed and used. Efficiency is an output-to-input ratio, such as monthly manufacturing output divided by the personnel hours used to produce that output. Effectiveness is the relationship of outputs to some standard or expectation, such as monthly manufacturing output expressed as a percentage of the goal for that month.

Efficiency measures have the advantage of being easier to measure and standardize across organizations, industries, and nations (Norman & Bahiri, 1972). When we hear that productivity growth in the United States has declined over the last 20 years (e.g., American Productivity Center, 1981), it is an efficiency ratio that is being quoted (i.e., price deflated gross national product divided by worker hours). Efficiency also has the advantage that it can be tied to profitability (Kendrick, 1984; Kopelman, 1986). Effectiveness is a much broader concept because it can include other factors, such as standards, objectives of the organization, expectations of interested parties (e.g., shareholders, regulatory agencies, and customers), and the viability of the organization relative to its competition. In addition, effectiveness can more readily include measures of quality. One of the biggest criticisms of the efficiency approach is that generating large amounts of output at low cost is dysfunctional if the output does not meet the objectives of the organization; a warehouse full of efficiently made products that no one wants to buy is not good. One of the biggest criticisms of the effectiveness approach is that meeting organizational goals without considering the resources expended is not in the best interests of the organization either. Proponents of the effectiveness concept argue that as complexity and ambiguity of the work increase, effectiveness measures become more important than efficiency measures (Balk, 1975).

Obviously, both efficiency and effectiveness have advantages. In fact, many authors define productivity as a combination of both (Balk, 1975; Coulter, 1979; Hanes & Kriebel, 1971; Kopelman, 1986; Mali, 1978; National Center for Productivity and Quality of Working Life, 1978; Sibson, 1976; Tuttle, 1981).

The next issue deals with what measures should be included in the measurement of productivity. Clearly, the different perspectives, such as the economist's and the accountant's, have implications for what measures are included, as does the organizational model used (Steers, 1975). There are, however, a variety of other possibilities. Campbell (1977), for example, listed 30 types of measures that have been used. These include the expected measures such as effectiveness, efficiency, profit, and quality as well as measures such as absenteeism, growth, morale, control, internalization of organizational goals, evaluations by external entities, and stability. As another example, Seashore and Yuchtman (1967) reported a factor analysis of organizational productivity scores for insurance agencies. They identified 10 factors, many of which were quite different from those listed by Campbell. They included new member productivity, youthfulness of members, business mix, manpower growth, and market penetration. These measures that might be included in a productivity measurement system clearly show that there is no one set of measures that defines productivity. The diversity of possible measures must be considered in the design of a productivity measurement system.

Most authors do agree on one issue about what measures to include: A productivity measurement system should include all important aspects of the organization's work (e.g., J. P. Campbell, 1977; Muckler, 1982; Tuttle et al., 1985). If the system is not complete, it could easily encourage neglect of the organizational objectives that are not part of the measurement system. In such a situation, the actual overall effectiveness of the organization would suffer (Duerr, 1974).

A final issue concerning productivity measurement is the use of an overall index of productivity. The use of a single index is very important both for motivational and for informational reasons. A single index is motivational because it provides the members of the unit with a measure of improvement or decrement. A single index is useful for informational purposes because it conveniently summarizes a large number of pieces of information on organizational functioning that could otherwise be very difficult to assimilate and use for making decisions. Many approaches to measuring productivity use a single index or argue for its value (e.g., Felix & Riggs, 1983; Joint Financial Management Improvement Program, 1976; Kim, 1980; Peeples, 1978; Pritchard et al., 1986, 1987a; Tuttle & Weaver, 1986a, 1986b; Tuttle et al., 1985).

Conclusions From the Productivity Literature

In summary, there are many approaches, perspectives, and issues relevant to productivity measurement. It is tempting to ask what is the best way to conceptualize productivity. However, we believe that this is the wrong question. Instead, the perspective, level of analysis, model, and inclusion of different measures should depend on what question one wishes to answer. Efficiency and effectiveness approaches both have their place, as do the different perspectives and models. How one resolves some of the other issues, such as use of a single index and what to measure, depends on the circumstances.

Put another way, the best way to measure productivity depends on the purpose of the measurement. Gauging the health of the national economy, identifying how an organization is doing relative to its industry competitors, and attempting to improve the productivity of the parts of the organization require different ways of measuring productivity.

Our Approach To Organizational Productivity

This argument suggests that to define our approach to productivity in the ProMES methodology, we must first identify the purpose of measuring it. In the simplest terms, our purpose in measuring productivity is to be able to increase it within a given organization or part of that organization. Thus, our perspective is closest to that of the industrial and organizational psychology
perspective. In taking this perspective, we assume that the behavior of the individuals in the organization has a meaningful effect on productivity. That is, productivity is not just determined by factors such as the organization’s environment, strategic decision making, and the organization’s technology. Furthermore, because our orientation is toward improving productivity, we take the position that productivity must focus on the things that the personnel in the unit or organization can control.

The mechanism by which this increase would occur is primarily a motivational one. Increased motivation means that personnel would exert more effort and be more persistent in their efforts. Their efficiency would increase because their efforts would be more directly related to organizational objectives, and they would cooperate more effectively to meet those objectives. They would improve their individual and group work strategies and would use their own and others’ time and efforts with less waste.

We also take a position that productivity is a group-level concept. Our position is that the concept of productivity should explicitly acknowledge that aggregate individual performance is not the same as productivity. Productivity includes the idea of a complex network of interrelationships at the intragroup, intergroup, and even interorganizational levels (Steiner, 1972). This conceptualization precludes including the individual level of analysis in the concept of productivity.

The model of organizations we accept is a composite of the goal-oriented, the natural systems, and the multiple constituency models. Our position is that all organizations have survival as a primary aim. However, beyond this overriding aim, they also have specific, identifiable goals that are stable enough to guide action. We agree with the natural-systems and multiple-constituency approaches that it is inappropriate to think that these goals are totally the product of rational decision making and are the sole determinants of organizational actions. These goals and objectives are frequently determined after considerable activity has taken place rather than before (Simon, 1964; Starbuck, 1965; Weick, 1969). In addition, the determination of objectives is a developmental, evolutionary, and highly political process that is less than totally rational, and objectives sometimes need to be set now for an unknowable future (Pennings & Goodman, 1977; Pfeffer, 1977; Pfeffer & Salancik, 1977, 1978). Finally, objectives are indeed the result of a process of negotiation of different constituencies with different needs and varying influence (Pennings & Goodman, 1977). However, with all these complexities, we take the position that for the practical purpose of measuring productivity, goals can be identified that are accurate, relatively stable indicators of the organization’s objectives as determined by the dominant organizational coalition (typically, top management).

We believe that a productivity measurement system should have the capacity for including both efficiency and effectiveness measures. There are so many compelling arguments for each that choosing one to the exclusion of the other is not possible. Although the system should be able to accommodate both, the choice of using either or both should depend on the particular situation. For example, if the unit has no control over inputs, efficiency is less meaningful than effectiveness. In the majority of cases, the appropriate approach is to first consider productivity as effectiveness rather than efficiency. We take this position for three reasons. First, effectiveness, with its orientation towards goal attainment, is a broader definition of productivity because it results in a measurement system that expresses productivity in terms of how good that productivity is. In contrast, an efficiency approach does not carry with it evaluative information on what is a good or bad level of efficiency. The second reason for adopting the effectiveness approach is that with this approach it is easier to generate a measurement system that can combine all aspects of the organization’s productivity into a single measure. The final reason is that the system we are proposing makes it possible to readily combine inputs with the effectiveness measure(s). Once an overall effectiveness measure is developed, it is a fairly simple matter to divide that effectiveness measure by inputs to get a system that combines the best aspects of both effectiveness and efficiency.

Given our purpose of increasing productivity, it is critical that the measurement system be complete so that increases in measured aspects of the work are not made at the expense of equally important but unmeasured aspects: What you measure is what you get. If the measurement includes all important aspects of the work, nothing important will be ignored. Finally, the individual measures should be combined into an overall measure of productivity for both motivational and informational purposes.

One of the objectives of this research program was to develop and evaluate a new approach to measuring productivity based on the foregoing conceptualization. With this background in mind, we now turn to the specific objectives of the research. These can be expressed as a series of questions.

1. Can this new approach to measuring productivity (to be described in detail later) be effectively developed in a functioning organization?
2. Will giving group-based feedback with the productivity measurement system increase productivity?
3. Will group-based goal setting and incentives increase productivity over feedback?
4. Will using group-based feedback, goal setting, and incentives change work attitudes?
5. Will any treatment effects continue after the departure of the research team?

Method

Site

The project was conducted with five organizational units at an Air Force base in the southwest United States. One, a maintenance section called Communications and Navigation (Comm/Nav), repaired a variety of electronic equipment used for aircraft communication and navigation. The number of personnel in Comm/Nav ranged from 29 to 35 during the course of the project. The other four sections together made up the Materiel Storage and Distribution Branch (MS&D). The four sections were Receiving, Storage and Issue, Pickup and Delivery, and Inspection. The MS&D branch was essentially the base warehouse. Property was delivered to the warehouse and checked in by the Receiving section. Storage and Issue shipped the property and retrieved it as it was ordered by sections on the base. The Delivery section delivered the property to sections on base that had ordered it. Inspection made sure the property was in good condition and that regulations concerning property packaging, storage, and identification were being followed. The number of personnel in the MS&D branch ranged from 47 to 54 during
the course of the project, with Receiving averaging 15, Storage and Issue, 15, Pickup and Delivery, 13, and Inspection, 7.

Description of the Productivity Measurement System

The theoretical background for the ProMES approach to organizational productivity stems from the theory of organizational behavior presented by Naylor, Pritchard, and Ilen (1980). In essence, we used the Naylor, Pritchard, and Ilen conceptual approach to roles and extended its application from individuals to organizational sections. This application led to the development of a number of unique features for a productivity measurement system. A second source for the development of our approach was the work of Tuttle (1981; Tuttle & Weaver, 1986a, 1986b; Tuttle et al., 1985). In this work Tuttle and his colleagues developed an approach to measuring productivity that included methods of going from what we call products to obtaining objective indicators of how well these products were being done. He used a variety of group techniques to develop products and indicators, some of which we also used.

The technique used to generate the ProMES is described in Pritchard et al. (1987a) and in more detail in Pritchard et al. (1986); the process will be summarized here. It consists of four distinct steps.

The first step is to identify products. Every organization has a set of activities that it is expected to do. These activities result in a set of what Naylor, Pritchard, and Ilen (1980) call products. Products can be thought of as the set of objectives that the organization is expected to accomplish; the productivity of the organization is a function of how effectively it generates these products. The first step in developing the productivity measurement system is to identify these products.

A series of meetings is held with unit supervision and representative incumbents to discuss and identify these products. These meetings continue until the group reaches consensus on the list of products. On the basis of our experience this would typically result in 3 to 6 products. In the example of the Comm/Nav unit, products would include meeting repair demand in a timely fashion, doing high quality repair, and so on.

The second step of the process is to develop what we call indicators of the products. An indicator of a product is a measure of how well the organization is generating the product in question. The indicators are developed in meetings by the same unit personnel who developed the products. They are asked to think of what would best represent how well they are producing their products. There may be only one indicator for a given product, or there may be more than one. Some indicators will already be available; some will have to be developed. The indicators are discussed and refined by supervision and representative incumbents until consensus is reached. In the Comm/Nav example, an indicator of meeting repair demand might be the number of items repaired divided by the number brought in for repair during the month. An indicator for quality of repair might be the percentage of items repaired that were returned as nonfunctional immediately after installation.

Once the products and indicators are identified and approved, the next step is to establish the contingencies. The term contingency should not be confused with the behaviorist use of the term to mean the relationship between a behavior and a reinforcer. In contrast, we use the term as it is used by Naylor et al. (1980): to refer to the idea that the level of evaluation of an outcome is contingent on the amount of that outcome. A contingency is the relation between the amount of the indicator and the effectiveness of that amount of the indicator. Figure 1 presents an example of a contingency. The upper half of the figure shows the general form of a contingency. The horizontal axis is the amount of the indicator. It ranges from the worst feasible level that the indicator could have to the best possible level. On the vertical axis of the figure is the effectiveness value of the various levels of the indicator. It ranges from +100 (maximum effectiveness) to -100 (minimum effectiveness). It also has a zero point, which is defined as the expected or neutral level of effectiveness. That is, the zero point is neither positive nor negative. For this example we have chosen the indicator for return rate: the percentage of items that did not function immediately after repair. Assume that the people in the organization indicate that the best return rate possible is 2%. They see it as impossible to do better than 2% because the electronic components they use for repairs can work properly when installed and checked, but a small percentage fail almost immediately when put into use. Assume also that they say that the worst possible return rate is 20%.

Once the best and worst possible levels of output have been established by the organizational personnel, the next task is to identify the actual function that relates the amount of the indicator to effectiveness. We start by determining the zero point. That is, what level of the indicator is the expected level, or the level that is neither especially good nor especially bad. Once this is established, a point is placed on the figure at the intersection of the zero point of the vertical axis and the level of neutral point on the horizontal axis. For example, if the neutral point was identified as a return rate of 10%, it would be so indicated as shown in the lower half of Figure 1.

Next, the effectiveness level of the maximum and minimum indicator levels would be established. The first step is to list the maximums for each of the indicators. The group of incumbents and supervisors is then asked to rank order these maximums in terms of the contribution of each to the overall effectiveness of the unit. The group discusses this and a consensus is reached. The maximum with the highest importance rank is then given an effectiveness value of +100, and the group is asked to rate the other maximums as percentages of the +100 maximum. For example, if the maximum of a given indicator was only half as important to the effectiveness of the unit as the most important maximum, they would give it a value of +50. An analogous process is done for the minimum values of each indicator, except the most important (worst)

![Figure 1. Example contingency.](image-url)
minimum is not constrained to be a value of \(-100\). It is given the value that the group feels is appropriate.

Once the zero points are identified and the effectiveness values of the maximums and minimums established, the remainder of the points in the function are developed by the group. Group discussion is continued until a consensus is reached.

For example, assume that the personnel in the organization said that to be at the minimum return rate would correspond to an effectiveness value of \(-80\), and to be at the maximum would be a \(+70\). After the other points on the function were identified, this might result in a contingency similar to that in the lower half of Figure 1. It indicates that going above the neutral point of a 10% return rate is positive, but this increase is not linear. Once they get to a return rate of 6%, further increases do not represent as great an increase in effectiveness. Likewise, at the low end, once the return rate gets as bad as 14%, they are doing very badly and further decreases are proportionally not as bad.

This process would be completed with each of the indicators, and once they were all scaled and reviewed for accuracy, the contingency set would be complete.

Two things are particularly noteworthy about the contingencies. The first is that the overall slope of the function expresses the relative importance of the indicator. The steeper the overall slope, the more important the indicator is to overall effectiveness. If meeting repair demand was more important to the unit than maintaining a low return rate, the contingency for meeting repair demand would be steeper than the contingency for return rate. The repair demand contingency might, for example, range from \(+100\) to \(-90\). Thus, the contingencies reflect differential importance. Aspects of the organization's work that are more important get steeper contingencies than aspects that are less important.

The second noteworthy aspect of the contingencies is that they can be nonlinear. This is necessary to accurately reflect the realities of an organization's functioning. In many cases, the relationships that actually exist are simply not linear. A linear relationship would mean that to improve a given amount at the low end of the measure is as good as improving that same amount at the high end. It would be very common, for example, for improvements in the middle range of an indicator to result in large improvements in productivity, but at the high end a point of diminishing returns is realized. Once the organization gets to a fairly high level of productivity on one aspect of the work, it is frequently better to work on improving something that they are not doing as well, than to continue to improve something that is already at a high level. For example, if the repair shop was operating with a very low return rate it might be better to improve its training than to further improve its return rate. Thus, even though return rate is more important overall than training, if return rate is good, improving a low degree of training can become more important to the overall effectiveness of the organization.

Once the contingencies are completed and approved by management, the last step is to put the system together. This begins with collecting the indicator data for a given period of time. Assume that the time period selected was a month. The data for the different indicators would be collected at the end of the month and, on the basis of the contingencies, effectiveness scores would be determined for each indicator by calculating the effectiveness for that level of the indicator. For example, if the maintenance unit had a return rate of 6% in a given month, examining the contingency in the lower half of Figure 1 indicates that such a return rate is associated with an effectiveness score of \(+60\). This value is determined by examining the contingency and noting that a 6% return rate on the horizontal axis is associated with an effectiveness value of \(+60\) on the vertical axis. Continuing this process gives an effectiveness value for each indicator. Once the effectiveness values for each indicator are determined, they can be summed to get the overall effectiveness of the unit.

These effectiveness scores have a distinct meaning. A score of zero would mean that organizational personnel were just meeting expectations; their productivity was neither particularly good nor bad. As the score became positive, they were exceeding expectations. The more positive the score, the more they were exceeding expectations. As the score became negative, they were below expectations.

This ability to simply sum effectiveness scores is a major feature of the system. Because the contingencies reflect the relative importance and the nonlinearity of the indicators, these factors are already incorporated in the system; thus, a simple summing does indeed reflect the overall effectiveness of the unit.

**Evaluation of the System**

It is clear that this approach to measuring productivity is heavily based on pooled judgments and thus incorporates subjectivity. Subjectivity is present in the listing of the products and indicators, and especially in the ratings that are used in the contingencies. Traditionally, subjectivity connotes lack of accuracy, unverifiability, and results of poorer quality than those obtained with "objective" data. Our position is that such subjectivity is not only acceptable, it is desirable in dealing with organizational productivity. The reason for this is that whether a given level of organizational output is high or low productivity is a matter of policy. The elements of the system are statements of this policy. Products, indicators, and contingencies indicate what is important to the functioning of the unit (and therefore by omission what is not important), the level of output that is expected (the zero point), how good other levels of output are, and the relative importance of different types of functions for the unit. This is policy, and policy is by nature a matter of judgment and thus inherently subjective. Our approach offers a way of deriving policy. It represents a way of reducing ambiguity in policy by formally discussing it, quantifying it, and subjecting it to formal review and approval by the management of the organization. It can also be a way to develop an explicit policy when one does not already exist.

Although there is no "objective" criterion against which to evaluate the productivity measurement system, it is critical to ensure that the policy of the decision makers is correctly reflected in the productivity measurement system. One aspect of this is that the listing of products and indicators must be complete. If there are important products or indicators omitted, the system can easily produce a situation where those things that are measured are attended to, and those that are not measured are somewhat ignored. This uneven attention to important functions can result in very dysfunctional consequences to the organization.

A second aspect of ensuring that the system accurately reflects policy is the degree to which the system accurately indicates what the unit should be doing. This means that the products, indicators, and contingencies that are used in the system must be consistent with policy and must be correctly scaled on effectiveness.

Both of these aspects of accuracy are dealt with in the development of the system by having a clear process of approval of the system at higher levels of the organization. This approval process is made clear from the start; at the beginning of the development of the system, all participants are told that incumbents and supervisors will develop the products and indicators. Once these are finalized, they are presented to higher management for approval. Once higher management has approved the products and indicators, the supervisory groups develop the contingencies and these are also formally approved by higher management.

**Features of the System**

The ProMES has a number of unique features. First, it allows for a clear statement of exactly what should be improved first, second, and so on to optimally improve overall productivity. Such a rank ordered set of priorities can be derived directly from the contingencies in the system. Second, the system allows for aggregating the productivity measurement of different units into a larger organizational unit. For example, if a department is composed of five separate units, the system devel-
ops a measure for each unit and allows the combination of those five measures into a single measure for the entire department. In most productivity measurement systems this is not possible because the measurement system varies from unit to unit. An advantage of the ProMES is that it is possible to do such across-unit aggregation. The method for doing this essentially amounts to a very simple additional scaling step in the development of the contingencies. These and other features of the productivity measurement system are explained in detail in the articles describing the productivity measurement system (Pritchard et al. 1986, 1987a).

Another unique feature of the system is that it allows the use of both effectiveness and efficiency approaches to productivity. In the previous example and in the actual research, the effectiveness approach was used. However, it would be a simple matter to divide the overall effectiveness score by inputs. These could be limited to labor inputs (e.g., number of personnel, costs of personnel), inputs that were controllable by the unit, or total inputs. This type of measure is essentially a cost-effectiveness measure. A second way to include inputs would be in the actual indicators. Percentage of repair demand met could be divided by the total personnel hours used. If appropriate, some indicators could be divided by inputs while some were not.

It is interesting to speculate on the relationship between effectiveness and efficiency measures. One would expect that overall effectiveness measures would be fairly highly related to overall efficiency over time. This would be expected not only because they are related conceptually, but also on a purely psychometric basis. Because effectiveness is output relative to expectations or goals and efficiency is output relative to inputs, the common measure of outputs should produce positive relationships between the measures. In fact, if expectations and inputs are constant, there should be a very strong relationship between effectiveness and efficiency over time. The relationship between the two overall measures should decrease as there were decreases in the relationships between (a) changes in inputs and changes in expectations, (b) changes in inputs and changes in output, and (c) changes in outputs and changes in expectations. One would expect that in most cases all these relationships would be positive, producing fairly strong positive relationships between overall effectiveness and overall efficiency. For example, one would expect that with high levels of inputs, expectations of output would be higher. Thus, there should be a positive relationship over time between a measure of overall effectiveness and overall efficiency. However, because the relationships among the three components can vary widely depending on the organization, the overall effectiveness-efficiency relationship is difficult to predict.

**Treatments**

Once the productivity measurement system was completed, the next phase involved instituting the interventions of feedback, goal setting, and incentives. To do this, a baseline was first established. This was a period of 8 months for MS&D and 9 months for Comm/Nav where indicator data were collected, but no feedback was given to the units.

The design of the project consisted of all five of the units developing first the productivity measurement system, followed by the baseline. Next, feedback was given to each unit for 5 months. Goal setting was then added to feedback for each unit and continued for another 5 months. Finally, incentives were added to feedback and goal setting in each unit for 5 months.

**Feedback.** Productivity feedback consisted of formal computer-generated reports that were given monthly to the personnel in each unit. To develop these reports we first produced examples of what the basic report would look like and asked supervision how best to present the material for maximum clarity. We also proposed some other information that they might find useful and asked them for their suggestions on things to be included. This was discussed, and after several revisions, a final version of the productivity feedback report was developed.

The first part of the resulting report was the basic productivity data. It showed the products and indicators, the indicator data for that month, the effectiveness score associated with that level of each indicator, and overall effectiveness for the unit. The report also included the indicator data and effectiveness scores for both the previous month and the current month, and the changes in effectiveness from the last month to the current month. This part of the report was requested by unit personnel to aid them in diagnosing areas where they were increasing or decreasing in productivity.

Another feature of the report was information comparing the productivity of different units. The system allows one to directly compare the productivity of very different units. This feature was very important to supervision and management of MS&D because it allowed them to compare the four sections of the branch. To make this comparison, we first determined the maximum possible overall effectiveness for each section by determining the effectiveness values for the maximum possible value of each indicator and summing these values. This sum represented the effectiveness value that would occur if the unit was doing as well as possible on every aspect of its work: its maximum possible overall effectiveness. Recall that these maximums were developed by consensus among the supervision of the units, and were discussed and approved by management. Thus, they should represent realistic maximums, and the effectiveness scores represent the value of the maximum contribution each of the units could make to the organization.

Once the maximum possible effectiveness was calculated, the actual monthly overall effectiveness score for each section was expressed as a percentage of maximum possible effectiveness. This percentage of maximum effectiveness was the measure by which each unit was compared with the others. This data was put in the monthly feedback report for each section of MS&D.

The feedback report was generated each month for each of the five units. It was presented within 3 working days after the end of the month and was given to the chain of command from section supervision to upper-level management. It was also posted in the working area of each section so incumbents could review it. In addition, graphs were posted in the work area and updated each month, one for overall effectiveness and one for each indicator. These graphs allowed unit personnel to easily see changes in effectiveness over time. As one might imagine, both the feedback report and these graphs generated considerable interest when they were posted each month.

Once the feedback reports were circulated, a meeting was held with incumbents, supervisors, and for MS&D, a representative from management to review the feedback report for the month. In this meeting, the report was reviewed. Areas of improvement were noted and areas of decrease discussed. Reasons for the improvements or decreases were considered, and any long-range trends were noted. This meeting also served as a basis for planning priorities for the next month and altering work strategies to improve productivity.

**Goal setting.** After 5 months of feedback, each of the five units began to set productivity goals in addition to receiving the feedback reports. The first step in implementation was to train the personnel in how to set goals. The nature and process of goal setting was explained to them, and the importance of participation was stressed. They were told the importance of setting a difficult but attainable goal. Prior to the start of the goal-setting program, a meeting was held with supervisors and incumbents to explain the nature of the program and how it would work. They were told to discuss the goal among themselves for a few days, after which the first goal setting sessions were held. These were attended by members of the research team to facilitate the process and especially to ensure participation and to encourage them to set a challenging but reasonable goal.

One issue that became quite significant was whether the goals would be "reportable" or "nonreportable." Unit personnel were reluctant to set goals that represented difficult but attainable goals if the goals were reported to higher management. They felt that they would be held to whatever goals they set and would receive negative evaluations if the goals were not met, even if they had actually increased their productivity.
ity. They felt that public or reportable goals would be set at a level that would be low enough to ensure that the unit would exceed them. Such goals would have little motivational force, and there is no reason to expect that a goal set lower than current productivity would have a positive effect on productivity. Thus, the goal-setting system was designed to be nonreportable; incumbents and supervisors set the goals, and members of the unit knew what the goals were, but the goals were not communicated to management beyond branch supervision. Upper management agreed to these nonreportable goals. They felt that if productivity continued to be as high as it had been during feedback, or if it improved, they did not need to know the unit goals that helped produce such results.

Comm/Nav and each of the four units of MS&D set goals for themselves. The goals were set in terms of overall effectiveness for the following month. They were set jointly by incumbents and supervisors. When the month-end feedback report was produced, the unit personnel noted whether they had attained the goal or not and discussed reasons why the goal was or was not met. At this time the goal was reviewed and reset for the following month.

Incentives. After 5 months of feedback and goal setting, incentives were added as the final treatment. Time off from work was chosen as the incentive following considerable discussion and input from samples of incumbents up to senior management. Specifically, unit personnel were given a half day or a full day off if the unit achieved the level of productivity needed to qualify for the incentive. This incentive seemed powerful for the personnel involved and was feasible to implement.

In addition to the incentive used, there were several other issues to be dealt with in designing the incentive system. One was the number of levels of productivity that would be defined as resulting in different incentive levels. At one extreme is a system in which there is a single level of productivity that, if reached, leads to an incentive. At the other extreme is a piece-rate-type system in which each increase in productivity leads to an increase in the amount of incentive. The advantages of the single-level approach are that it is simple to administer and can be used with any type of incentive. The advantage of the multiple-level approach is that no matter how high the unit's productivity is, there is still an incentive to increase further. With the single-level approach, once the level of productivity that produces the incentive is reached, there is no further incentive to increase.

Another design issue concerned asking the units for higher and higher levels of productivity. At the time the incentive system was being designed, all of the units had greatly increased their productivity. The system encouraged them to increase their productivity during feedback and again during goal setting. If the incentive system required even more improvement before any incentives were awarded, we felt there was a possibility of a negative reaction to the program.

There was also an issue of setting the incentive levels so that they were equitable across units. For Comm/Nav this was not a problem because they had little contact with MS&D. However, the four sections of MS&D were in constant contact with one another, and designing a system so that one section had a much easier time obtaining the time off than another would have led to problems of inequity.

Finally, there was the issue of whether to use branch-level incentives in addition to the section-level incentives in MS&D. The four units of MS&D had to cooperate extensively to get the work done. It would have been possible for a given section to spend time only on activities that would improve their own effectiveness and not devote efforts to activities that would help the other sections do their jobs. If an incentive system is built in such a situation with no mechanism for fostering cooperation between units, problems could potentially result.

The incentive system was designed with these issues in mind. The system used two levels of the incentive, a half day off and a full day off. Comm/Nav received a half day off if their overall effectiveness exceeded the mean of the five immediately preceding months. If they exceeded this level by a meaningful amount as determined jointly by incumbents, supervision, and management, they received a full day off.

This full day off was awarded when overall effectiveness exceeded the mean of the last 5 months by 5%. Because Comm/Nav was very close to their maximum possible effectiveness by this time, this 5% increase was judged to be a substantial improvement. One problem with this system was that if productivity continued to increase, the mean of the last 5 months would continue to rise. This would mean that it would be more and more difficult for the unit to achieve the productivity level necessary for incentives. Eventually, continued increases in productivity would make it impossible to get any incentives. To deal with this problem, a maximum was set on the overall effectiveness level necessary to get the incentives. For the half day off, the maximum level was set at 85% of their highest possible overall effectiveness; for the full day off, the maximum was set at 94%. Thus, the unit received a half day off if their effectiveness was greater than the mean of the last 5 months or was greater than 85% of maximum possible effectiveness. The unit received a full day off if their effectiveness was 5% higher than the mean of the last 5 months or was greater than 94% of maximum possible effectiveness.

MS&D also had two levels of the incentive, but with a different structure. A given section received a half day off if its overall effectiveness exceeded the mean overall effectiveness of the 5 highest productivity months prior to the start of incentives. In practice, this meant that they would receive the incentive if their productivity exceeded the mean productivity under goal setting. There was also a branch-level incentive of an additional half day off that would be given to each of the four sections if the branch reached its incentive level on overall effectiveness. This productivity level for the branch to get the additional half day off was essentially the sum of the section-level goals. That is, the individual section incentives were given when the section's productivity continued at a very high level. To get the branch incentive, productivity had to be maintained at this high level for each section. This was done because several of the sections were already performing very near the maximum possible level, and asking them to go above this to get the full day off as in Comm/Nav was unreasonable. If a section reached its incentive level and the branch reached its level, that section received a full day off. If the section reached its level, but the branch did not, the section received a half day off. If the section did not meet its incentive level, but the branch did, the section received a half day off.

Thus, the system dealt with the issue of asking for higher and higher productivity in Comm/Nav by having one level of the incentive for continuing to maintain previous high levels of productivity, and a second level of incentive for exceeding that productivity. For MS&D, this issue was dealt with by requiring that a section maintain its high level of productivity for the half day off, and that all sections maintain high productivity for the full day off. The branch-level incentive used in MS&D was structured so that sections were encouraged to cooperate. To get the greater incentive, the entire branch had to do well. A section could have a bad month and still get a half day off if they could help the rest of the sections pull up the branch total. Another possibility was that one section was doing extremely well would help another section so that both would get time off.

The issue of equity across sections in MS&D was dealt with by first explaining the issue of equity to the personnel and then having incumbents, supervision, and management of the different sections jointly determine what the levels of productivity should be to obtain incentives in each of the sections.

During the incentive treatment, the units continued to set their own performance goals and review them each month. It was felt that after the incentive treatment started they might simply set their goals equal to the level necessary to receive the time off. Thus, continuing goal setting might be unnecessary. However, the goal setting was continued so that the individual units could set different goals if they so desired and so that the integrity of the experimental design was maintained. Specifically, by allowing units to continue their own goal setting, we could assess the incremental effects of incentives when added to goal setting. If goal setting was discontinued, assessment of this incremental effect would not have been possible.
Results

The results of the project will be presented in five sections: (a) development of the productivity measurement system; (b) effects on productivity; (c) results for the feedback, goal setting, and incentive interventions; (d) attitude data; and (e) effects after the departure of the research team. Additional details about the results can be found in Pritchard et al. (1986, 1987b).

Development of the Productivity Measurement System

There were a number of results pertinent to the development of the ProMES, including qualitative data, results of reliability and validity checks on the productivity measurement system, and subjective reactions to the system. These findings are discussed in detail in Pritchard et al. (1986, 1987a), but can be summarized as follows.

1. Attitudes toward the development of the system went from mixed at the start of the project to very positive at the end of system development.
2. Personnel were able to develop the system. They had little difficulty with identifying products, had considerable difficulty generating quality indicators, and had little difficulty developing contingencies. They seemed to get particularly involved in the development of the contingencies, reporting to us that the contingencies really captured policy by clearly showing how the unit was supposed to operate.
3. One of the units had both a day shift and a night shift and contingencies were developed separately with the two shifts. Interjudge reliability of the contingencies was very high.
4. All of the contingencies were nonlinear.
5. Validity data collected on the system were positive.
6. Questionnaire reactions to the system on such factors as liking of the system, perceptions of system worth, degree to which the system helped productivity, and so on, were very positive.
7. In all the units, managers asked that the system be installed in other units in that part of the organization.

Effects on Productivity

Once the system was developed and a baseline was established, the system was used to generate feedback. Next, goal setting was added to the feedback, and then incentives were added. The overall effectiveness scores over the five units are shown in Figure 2. This figure is a plot of the mean of the overall effectiveness of the five sections by time, showing the baseline and the three treatment periods. As the figure indicates, overall effectiveness increased substantially over the baseline.

In order to calculate a single index of change in productivity, the percentage gain from the baseline compared with the maximum possible gain was calculated. That is, mean overall effectiveness was calculated from the baseline. The difference between this value and the maximum possible overall effectiveness the section could obtain was calculated. This maximum possible effectiveness was determined by calculating the overall effectiveness score if the unit produced at the maximum possible value on each of the indicators. The gain in overall effectiveness from baseline to each treatment was then calculated, and expressed as the percentage of maximum possible gain. For example, assume that the baseline mean was 400 and the maximum possible overall effectiveness was 800. If the mean overall effectiveness during feedback was 600, this would be a gain of 200. This is 50% of the maximum possible gain and would be the value reported. This approach to calculating change in some ways is conservative in the sense that the maximum possible increase is limited to 100%. The situation where the percentage of maximum gain measure would not be conservative is where the maximum possible gain was small, thus making small gains appear to be very large. This is not a problem here because the maximum possible gains from baseline were large.

With this approach, average increase over baseline was 50% for feedback, 75% for goal setting, and 76% for incentives. For Comm/Nav alone, the increases were 30% for feedback, 65% for goal setting, and 68% for incentives. For the four sections of MS&D, percentages of increase were 54%, 77%, and 79%. As will be discussed later, results for each of the four sections in MS&D were similar to these overall figures.

These results indicate a major increase in productivity. The effects were extremely large. However, before discussing them in more detail, some interpretation issues must be addressed.

Potential problems of interpretation. Before one can confidently attribute these results as effects attributable to the experimental interventions, several issues must be considered. The first is the possible presence of a Hawthorne effect, where productivity could have increased simply because the units were
singed out for the special treatment of being in a research project. Although such an effect is indeed possible, the project was designed to avoid possible contamination from such an effect. Specifically, the initial contact with the units during system development was quite intensive, and if a Hawthorne effect was going to occur, it should have occurred then. Because this contact started well before baseline, any productivity increase should have already occurred and would not contaminate the results. Thus, the presence of a Hawthorne effect cannot explain the productivity increases.

Another issue in the interpretation of the results is the number of personnel in each unit. Specifically, the increases in productivity that occurred during the treatments could be due to increases in the number of personnel in the units. For Comm/Nav the data collected were the total number of personnel in the shop. The mean number of personnel during baseline was 30.9. This figure increased slightly during feedback to 33.0, was 32.8 during goal setting, and dropped back to 31.0 during incentives. Because personnel levels during the period of highest productivity (the incentive treatment) were essentially equal to the level during baseline (30.9 vs. 31.0), we conclude that the increases in productivity were not caused by changes in number of personnel. In MS&D, the data collected were total number of personnel and the number of hours of overtime logged per month. Unlike Comm/Nav, MS&D routinely had considerable overtime. The mean number of personnel in MS&D was 51.8 for baseline, 53.7 for feedback, 48.4 for goal setting, and 49.2 for incentives. Thus, the overall number of personnel decreased over the period of the treatments. The overtime data indicate that number of hours of overtime decreased during the time productivity was going up. Overtime went from a mean of 1,348 hr per month during baseline to 892 hr during feedback, 404 hr during goal setting, and 416 hr during incentives. Thus, by the end of the treatments, overtime was less than one third of what it was during baseline. These data indicate that by the end of the project, the productivity gains that had occurred were achieved with no increase in number of personnel in Comm/Nav and a decrease in personnel in MS&D.

A final point in the interpretation of the results is that there could have been changes occurring in the larger organizations of which the five experimental units were a part that were causing general increases in productivity for all units. To explore this possibility, data were collected on several comparison groups that were similar to the experimental groups in the type of work they did and the larger organization of which they were a part.

Comparison data for Comm/Nav consisted of 10 measures of productivity from seven other maintenance units that were part of the same larger maintenance organization. These units were similar to Comm/Nav in that they repaired other types of aircraft equipment. These units had no intervention from the research team. The data were collected from the baseline through the end of the treatments and the mean of these 10 measures is an index of the overall productivity of these comparison units. This mean was 317 during baseline, it dropped to 295 during feedback, and rose to 377 and 365 during goal setting and incentives, respectively. These values represent a decrease during feedback of 6.9% and increases of 18.9% and 15.1% during goal setting and incentives, respectively.

These results show that the comparison groups decreased somewhat in productivity during the Comm/Nav feedback period and increased thereafter. This would suggest that the productivity increase during feedback in Comm/Nav was not due to wider organizational changes affecting all units in that part of the organization. Units similar to Comm/Nav decreased during this period. Furthermore, the increases during the Comm/Nav goal-setting and incentives periods were not present across all comparison group units. These increases were primarily brought about by 2 of the 10 comparison measures showing fairly large increases and thus do not reflect a general trend across the measures.

There were four comparison measures reflecting overall functioning of the supply squadron (exclusive of MS&D) and two other units outside of Supply that were under the same general management as Supply. The mean of these four measures during baseline was 516, and was 512 for feedback, 511 for goal setting, and 518 for incentives. These means represent changes from baseline of less than 1%. Thus, there were no changes in productivity for the MS&D comparison groups.

Taken together, the comparison group data indicate that the effects on productivity that occurred in the experimental units cannot be explained because of wider organizational changes in productivity.

Taken as a whole, the doubtful impact of possible Hawthorne effects, the data on the number of personnel, and the comparison group data make a persuasive case that the changes in productivity that occurred were because of the experimental interventions. In addition, our subjective impressions and the reactions of the personnel in the units completely convinced us that this was the case.

Indicator data. As positive as the results on overall effectiveness are, it is also instructive to examine the effects on the indicators themselves. These results are first presented in Table 1, which shows the means of the three most important indicators by treatment for Comm/Nav and for the four sections of MS&D. For Comm/Nav, the three most important indicators were percentage of repair demand met, percentage of bounces (items that did not function upon installation), and number of units awaiting maintenance. In MS&D, the most important indicator for Receiving, Storage, and Issue, and Pickup and Delivery was the speed with which they handled "Priority 2" material. Priority 2 items were the most important items they dealt with, usually a part for an aircraft that could not fly without that item. The other two most important indicators for Receiving were a type of error that could be made (in-checking errors) and speed of handling Priority 4 material. Receiving did not handle Priority 3 material. The other two most important indicators for Storage and Issue were handling Priority 3 material and storing items that did not have readily determinable locations (clearing items off R36 list). Pickup and Delivery handled two types of Priority 2 material, and the other most important indicator was handling Priority 3 material. The most important indicator for Inspection was the number of aircraft parts that had not been inspected by the end of the day. The fewer the number of parts, the more effective the unit. The other two most important indicators were inspections to test whether items functioned properly and inspections of items that appeared damaged.

These results indicate that the effects of the system were very powerful. Inspection of the table indicates that major improvements in the indicators occurred.
Table 1
Indicator Means by Treatment

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Baseline</th>
<th>Feedback</th>
<th>Goals</th>
<th>Incentives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications &amp; Navigation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% demand met (+)</td>
<td>88.5</td>
<td>90.5</td>
<td>93.9</td>
<td>92.6</td>
</tr>
<tr>
<td>% bounces (-)</td>
<td>5.9</td>
<td>4.5</td>
<td>2.7</td>
<td>2.5</td>
</tr>
<tr>
<td>No. units awaiting maintenance (-)</td>
<td>23.3</td>
<td>18.8</td>
<td>11.8</td>
<td>10.4</td>
</tr>
<tr>
<td>Material Storage &amp; Distribution (MS &amp; D) receiving</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority 2 receiving (in minutes) (-)</td>
<td>118.6</td>
<td>20.5</td>
<td>16.2</td>
<td>16.3</td>
</tr>
<tr>
<td>In-checking errors/1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority 4 receiving (hours) (-)</td>
<td>2.2</td>
<td>1.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>MS &amp; D storage and issue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority 2 issue (in minutes) (-)</td>
<td>23.4</td>
<td>5.3</td>
<td>6.5</td>
<td>4.3</td>
</tr>
<tr>
<td>Priority 3 issue (in minutes) (-)</td>
<td>60.8</td>
<td>13.4</td>
<td>9.5</td>
<td>8.9</td>
</tr>
<tr>
<td>Priority 4 issue (in minutes) (-)</td>
<td>117.7</td>
<td>22.3</td>
<td>15.8</td>
<td>20.9</td>
</tr>
<tr>
<td>% cleared off R36 list (+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS &amp; D pickup and delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority 2 issue (in minutes) (-)</td>
<td>89.1</td>
<td>96.5</td>
<td>98.9</td>
<td>98.3</td>
</tr>
<tr>
<td>Priority 3 issue (in minutes) (-)</td>
<td>64.0</td>
<td>35.3</td>
<td>27.4</td>
<td>27.1</td>
</tr>
<tr>
<td>Priority 4 issue (in minutes) (-)</td>
<td>98.3</td>
<td>42.0</td>
<td>33.1</td>
<td>34.2</td>
</tr>
<tr>
<td>Priority 3 (in minutes) (-)</td>
<td>76.5</td>
<td>52.0</td>
<td>35.9</td>
<td>35.9</td>
</tr>
<tr>
<td>MS &amp; D inspection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. aircraft parts left (-)</td>
<td>2.2</td>
<td>0.7</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>No. functional checks left (-)</td>
<td>0.6</td>
<td>0.1</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>No. suspect items left (-)</td>
<td>0.5</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note: A minus (-) after an indicator indicates that a smaller mean is higher productivity, a plus (+) after an indicator means that a larger mean is higher productivity.

Finally, the intercorrelations between the effectiveness scores for the indicators were examined. For each of the five units, correlations between the monthly effectiveness scores for the indicators for that unit were calculated over the baseline and three treatment periods. For example, for Comm/Nav with its 13 indicators, an intercorrelation matrix of the 13 variables (the effectiveness scores) was calculated using a sample size of 24 (9 months of baseline plus 15 months of treatments). The mean correlation of these effectiveness scores was .32 for Comm/Nav, .61 for Receiving, .22 for Storage and Issue, .28 for Pickup and Delivery, and .32 for Inspection.

Feedback, Goal Setting, and Incentive Results

Although the results we have been presenting so far deal with the effects of feedback, goal setting, and incentives, it is instructive to focus on them more directly. The results that directly compare the three treatments are presented in Figure 3. This figure is a plot of mean overall effectiveness by treatment for the four sections of MS&D and for Comm/Nav. (The means plotted in Figure 3 come from monthly data showing fairly smooth curves like those in Figure 2. Thus, they could be seen as exaggerating the differences between treatments.)

It is clear from this figure that a very consistent pattern resulted. There was a major effect attributable to feedback, a smaller but still large additional increase during goal setting, and either no increase or a very small additional increase during incentives. (Recall that goal setting should be considered as feedback + goal setting and incentives should be considered as feedback + goal setting + incentives.) It is tempting to conclude that feedback increased productivity, goal setting increased it considerably more, and incentives had little additional effect. However, we do not believe this is the correct interpretation. The continued effects of feedback and ceiling effects must be considered in the interpretation of the results. These issues will be presented in the Discussion section.

Effect sizes were calculated with the procedures from the Guzzo et al. (1985) meta-analysis. That is, the d statistic (Cohen, 1969) was used, whereby the mean difference between conditions is divided by the value of the pooled within-groups standard deviation as suggested by Hunter, Schmidt, and Jackson (1982). This pooled standard deviation is the square root of the sum of squares within the baseline monthly overall effectiveness scores plus the sum of squares within the treatment monthly overall effectiveness scores, divided by the degrees of freedom for baseline plus the degrees of freedom for the treatment. Next, the correction for sample size suggested by Hedges (1981) was made. This procedure results in effect sizes directly comparable to Guzzo et al.'s (1985) results. These results are presented in...
This difference in goal difficulty is reflected in actual goal attainment. The third column of the table presents mean percentage of goal attained. If the unit’s productivity equaled its goal, the value would be 100. If productivity exceeded the goal, the value would be above 100. The fourth column shows the number of months out of the 10 in which each unit achieved its goal. Both of these columns indicate that Storage and Issue personnel fairly consistently exceeded their goals, followed by Comm/Nav, with the rest of the units exceeding their goals less frequently.

The data indicate that all the units except Storage and Issue were generally setting goals that actually did represent increases in productivity. This may seem to be a surprising conclusion because there were several months when units set goals below their last month’s productivity. These lower goals were the result of some anticipated change in work load or resources for the upcoming month that would lower their productivity. For example, if Pickup and Delivery had a driver on leave for the month, they would lower the goal for that month.

The results also suggest that the goals that were set were fairly difficult ones in each unit except Storage and Issue. Although the percentage of increase of the goal over the past month’s productivity may seem low, this number included those goals that were set below the past month’s productivity because of work load or resource changes. In addition, all of the units were producing at near their maximums, especially during the last 5 to 8 months of goal setting. Thus, even a small increase in productivity was difficult to achieve.

Although Storage and Issue set substantially lower goals than did the other sections, this did not seem to have a negative effect on their productivity. Specifically, their percentage productivity increase over baseline during goal setting was 80% as compared with an increase of 77% for all four sections of MS&D combined. When incentives were added to goal setting, their increase was 81% over baseline, and all of MS&D showed an increase of 79%.

The incentive treatment lasted 5 months. During that time, the frequency of the units earning incentives varied considerably. Comm/Nav received time off least frequently. They received one half day off and one full day off over the first opportunities. Receiving and Storage and Issue each earned two half days and three full days off. Pickup and Delivery earned three half days and two full days. Inspection earned four full days and no half days.

Another point of interest relates to the way goal setting was set. This is described in Table 2. The table presents effects sizes relative to baseline. That is, for the feedback treatment, values in the table are the mean difference in overall effectiveness between baseline and feedback divided by the pooled within-groups standard deviation of baseline and feedback. The values for goal setting are the mean difference in overall effectiveness between baseline and goal setting divided by the pooled within-groups standard deviation of baseline and goal setting.

Effect sizes are presented for each of the five units separately as well as for the mean of the five units. In addition, results for MS&D overall are presented. These are the effect sizes for overall effectiveness scores for the entire MS&D branch. Finally, effect sizes are presented for the total overall effectiveness of the five units combined. In addition, the mean and range of effect sizes reported by Guzzo et al. (1985) in their meta-analysis of productivity enhancement techniques are included in the table.

It is clear that the effect sizes are very large. Even using the most conservative values for the effects over the five units (the means across the five units), these effect sizes are much larger than those found in Guzzo et al.’s (1985) meta-analysis. The feedback effects are 7 times as large as the mean feedback effect they report, the goal-setting effects are 6 times as large as the mean goal-setting effect they report, and the incentive effects are 8 times as large as the mean incentive effect they report. In fact, the effects are at least 3.5 times as large as the largest effect they report in any of the studies they review.

It is also instructive to examine what the units did in terms of goal setting and incentives. The goal setting data are presented in Table 3. Difficulty of the self-set goals varied considerably across the five units. This can be seen by the first column of the table, which indicates the number of goals, out of the total of 10 possible, where the month’s goal was set lower than productivity had been for the previous month. The values range from 0 for Receiving and Inspection to 8 for Storage and Issue. Goal difficulty can also be seen from the second column of the table, which presents the goal as a mean percentage of last month’s productivity. That is, if the unit set a goal that was, on average, 10% higher than their last month’s productivity, the tabled value would be 110. As can be seen, most of the units set goals that were higher than their last month’s productivity. The one exception was Storage and Issue, which set goals about 5% below productivity for the last month.

### Table 2

<table>
<thead>
<tr>
<th>Unit</th>
<th>Feedback</th>
<th>Goal setting</th>
<th>Incentives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications &amp; NAV</td>
<td>1.84</td>
<td>4.17</td>
<td>4.17</td>
</tr>
<tr>
<td>Receiving</td>
<td>2.73</td>
<td>5.64</td>
<td>5.74</td>
</tr>
<tr>
<td>Storage &amp; Issue</td>
<td>3.70</td>
<td>6.24</td>
<td>6.53</td>
</tr>
<tr>
<td>Pickup &amp; Deliver</td>
<td>1.75</td>
<td>3.76</td>
<td>4.06</td>
</tr>
<tr>
<td>Inspection</td>
<td>2.17</td>
<td>2.92</td>
<td>2.93</td>
</tr>
<tr>
<td>Mean of 5 units</td>
<td>2.44</td>
<td>4.54</td>
<td>4.68</td>
</tr>
<tr>
<td>MS &amp; D overall</td>
<td>3.42</td>
<td>9.11</td>
<td>10.43</td>
</tr>
<tr>
<td>Total of 5 units</td>
<td>3.77</td>
<td>9.37</td>
<td>11.03</td>
</tr>
<tr>
<td>Guzzo, Jette, &amp; Katzell (1985) meta-analysis effect sizes</td>
<td>M</td>
<td>.35</td>
<td>.75</td>
</tr>
<tr>
<td>Range</td>
<td>.08–.62</td>
<td>.57–.93</td>
<td>-.10–1.24</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Unit</th>
<th>No. of goals set below last month’s productivity</th>
<th>Mean % goal was of last month’s productivity</th>
<th>Mean % of goal attained</th>
<th>No. of times goal was attained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications &amp; NAV</td>
<td>2</td>
<td>100.9</td>
<td>102.5</td>
<td>6</td>
</tr>
<tr>
<td>Receiving</td>
<td>0</td>
<td>103.2</td>
<td>97.2</td>
<td>2</td>
</tr>
<tr>
<td>Storage &amp; Issue</td>
<td>8</td>
<td>94.7</td>
<td>106.8</td>
<td>8</td>
</tr>
<tr>
<td>Pickup &amp; Delivery</td>
<td>4</td>
<td>101.1</td>
<td>99.7</td>
<td>3</td>
</tr>
<tr>
<td>Inspection</td>
<td>0</td>
<td>102.9</td>
<td>99.4</td>
<td>3</td>
</tr>
</tbody>
</table>
conducted under the incentive system. Recall that goal setting was continued independent of incentives. Thus, there was a level of overall effectiveness established for the full and half days off and a level of overall effectiveness set by each unit as a goal. The question is whether goal setting was superseded by the incentive system so that the units set goals at a level equal to that necessary to obtain one of the incentives. Comparison of the goal data with the incentive levels indicates that there was no tendency for the interventions to converge. The goals that were set were independent of the incentive levels. Out of the 25 opportunities (5 units × 5 months), only 3 goals set by the units matched either of the incentive levels. In addition, observation of the goal-setting sessions suggested that the goals were being set independent of what levels were required for the incentives.

There was no identifiable pattern to the goals in relation to the incentive levels. Specifically, in 8 of the 25 possible occasions, the goals were set above the highest effectiveness level that was needed to earn a full day off. In 10 of the 25 occasions, the goals were set below the lowest effectiveness level that was needed for a half day off. For the remaining 7 occasions, the goals were set at or somewhere between the level needed for the half and full day off.

**Attitude Data**

Data were also collected on work attitudes. A questionnaire was administered to incumbents and first line supervisors in each of the units. It measured job satisfaction, reenlistment intentions, morale, role clarity, clarity of objectives, and evaluation clarity. Job satisfaction was measured by seven items adapted from the Minnesota Satisfaction Questionnaire (Weiss, Dawis, England, & Lofquist, 1967). The items for the morale scale were adapted from ISR instruments (Seashore, Lawler, Mirvis, & Cammann, 1983). Items for turnover intentions, clarity of objectives, and evaluation clarity were developed for this project. Items for individual role clarity were adapted from the Rizzo, House, and Lirtzman (1970) instrument.

The questionnaire measuring these scales was administered four times: during baseline and at the end of each of the three treatments. Internal consistency reliability was assessed using Cronbach's alpha (Nunnally, 1978), which was calculated for each scale on the basis of the data from the first administration. Internal consistencies were good for job satisfaction (.82), morale (.86), and role clarity (.87). They were somewhat lower for evaluation clarity (.76) and for clarity of objectives (.64), and quite low for the two-item turnover intention scale (.39).

The results by treatment are presented in Table 4. The means by treatment show that attitudes generally improved or stayed the same during the treatments. The data were analyzed to assess overall attitude differences between baseline and the mean of the three interventions. An independent-groups multivariate analysis of variance (MANOVA) was used where the six attitude variables were used as dependent variables. A repeated-measures analysis was not used because the number of individuals who remained in the units during all four administrations was small. Thus, the significance levels are conservative. The results indicated that attitudes under the three interventions were significantly higher than attitudes during baseline (Wilks's $\lambda = 900$, $df = 6$, $175$, $p = .005$).

The means for the four conditions were next analyzed with a one-way analysis of variance (ANOVA) for each of the six attitude variables, with the four administrations serving as a between-groups factor. As with the MANOVA, a repeated-measure analysis was not used because the number of individuals present in all four administrations was small. The means for each treatment are indicated, along with the Ns, the $p$ value for the between-administrations variable, and the error term for the between-administrations significance test (MSW).

The results indicated that the measures of turnover intentions, individual role clarity, and clarity of objectives showed no significant change, although the means for clarity of objectives became more positive, and the change was close to significant ($p = .08$). Job satisfaction, morale, and evaluation clarity became significantly more positive.

One general trend in the means is for the scales to be less positive during the incentive treatment. To test this, a MANOVA compared the means of the feedback and goal-setting treatments with the mean under incentives, using the six attitude measure as dependent variables. The difference was not significant (Wilks's $\lambda = .9735$, $df = 6$, $175$, $p = .576$). Similar nonsignificant results were obtained when attitudes under goal setting were compared with attitudes under incentives.

In summary, the overall results of the attitude data indicate that job attitudes under the treatment were as favorable or more favorable than before.

**Effects After the Departure of the Research Team**

Another aspect of the results is what happened after the researchers departed. When the 5-month incentive treatment was
over, the on-site responsibilities of the research team officially ended. Although we were on-site for a variety of purposes after this time, the units had no commitment to continue the system. These findings are discussed in Pritchard et al. (1987a, 1987b), but are summarized here. In each of the five units, the system was continued after the departure of the research team, and we were asked to implement the system in other units.

At the end of the formal incentives treatment, units were asked if they wished to modify the system. If they did want to, we would be there to assist if it became necessary. Comm/Nav and two sections of MS&D elected to make changes. In all three cases, the changes were to eliminate indicators from the system. The indicators that were removed were ones with very flat contingencies indicating they were not very important, were activities that the units were no longer going to perform, or were indicators that they felt were under such good control that they no longer needed to be measured.

A strength of the system is that it can accommodate changes readily. As changes occur in policies, procedures, or resources, changes will need to be made in the system. This can be done by eliminating indicators, redefining them, or altering the scaling of contingencies. Thus, the system can easily be altered to changing conditions. However, there is still a problem. After such changes are made, the new effectiveness scores are no longer comparable to the old scores. For example, if indicators are dropped, the same actual productivity will show up as lower overall effectiveness because some effectiveness points are lost due to the deleted indicators. This makes interpretation of effects over time difficult until a new "baseline" is established. This problem was dealt with by taking the revised system of indicators and contingencies constructed at the end of our involvement at the base and calculating overall effectiveness for several months prior to the revision. In the example of some indicators being dropped, it is a straightforward matter to go back to the indicator data from prior months and calculate what overall effectiveness would have been if those indicators had not been included. This becomes the new baseline, and the effectiveness scores after the change can be compared directly with this new baseline.

A final question to be considered is what has happened to productivity since the units have taken over the system themselves. The results are shown in Figure 4. This figure not only indicates the effects for the period after the units took over system operation, but also demonstrates the previous point about the ability of the system to generate a new baseline when changes in the system are made. Because both Comm/Nav and some sections of MS&D deleted indicators from the system, comparing overall effectiveness before the change to after the change is inappropriate. To deal with this, we recalculated the overall effectiveness scores back in time using the revised system to get comparability. In this case, all changes in the system were made the month after the 5 months of the incentive treatments, when our involvement in the interventions ended. To develop an appropriate comparison, we recalculated the overall effectiveness data for the incentive treatment. This adjusted overall effectiveness data is the overall effectiveness score that the units would have had if they had been under the revised system.

The mean overall adjusted effectiveness score under incentives for Comm/Nav was 519, and the mean for the next 10 months (when they had control of the system) was 520. Thus, they showed no decrease after the departure of the research team. The operation of the system has survived a change in the management above Comm/Nav that occurred after our departure and shows every indication of being a permanent part of the operation of the unit.

While it was operating, the system also did well in MS&D. It operated successfully without the research team for 5 months after our departure. During that time, productivity was just slightly below productivity during incentives. The data are plotted in Figure 4. The mean adjusted overall effectiveness during incentives was 1,857; the mean after our departure was 1,792.

However, the system was not continued in MS&D as it was in Comm/Nav. After the 5 months of operation, although much of the indicator data were collected, they were not put through the programs to produce the feedback reports, and thus feedback was not given to unit personnel. After several months of this, the units stopped trying to operate the system.

Discussion

To discuss the results of the effort, we will first consider the results pertinent to the research questions; we will then discuss issues pertinent to group-based interventions of productivity measurement, feedback, goal setting, and incentives.

Research Questions

The objectives of the research were presented earlier as a series of research questions. We will repeat these questions and discuss the results pertinent to each.

The first research question was whether the ProMES could be effectively developed in an actual organization. On the basis of the results reported in and Pritchard et al. (1986, 1987a), the answer to this question is a clear yes. The system appears to be a very effective way of measuring productivity. Its implementation is quite feasible, unit personnel were cooperative in developing the system, and it showed good psychometric properties. Finally, the system appeared successful in aggregating productivity measures across units so that an integrated system could be developed across the four sections making up the MS&D branch. This process is actually quite simple once the basic system is developed in each section (Pritchard et al., 1986, 1987a). The application of this aggregation to much larger and more complex organizational units seems quite feasible.

The second research question was whether using the group feedback resulting from the productivity measurement system would increase productivity. Feedback had a very strong effect on productivity. An average gain in productivity of 50% occurred across the five units during feedback. The most conservative estimate of the effect size of this difference is 2.44; the most liberal is 3.77. The effect sizes are several times larger than the largest effect size reported in Guzzo et al.'s (1985) review. This increase occurred with no change or with a decrease in manpower. Competing explanations for the positive results from the Hawthorne effect, increases in personnel, and increases brought about by other organizational changes can be ruled out.

The development of the system and its positive effects worked well on units that were quite different from one another. The units differed greatly in the nature of their work, with Comm/
Nav doing repair of sophisticated equipment and MS&D operating a large warehouse. The technologies were quite different between the two units, as well as among the four sections of MS&D. There were also great differences in the type of organizational structure and the work flow. The units varied considerably in size, and the personnel varied considerably in academic as well as in technical education. They also differed in initial level of performance, with MS&D acknowledged to be somewhat marginal in performance and Comm/Nav being perceived as fairly high. Yet with all of these differences, the system was developed and worked extremely well in each unit. This adds considerable support to the generalizability of the approach.

The third research question was whether goal setting and incentives would increase productivity over feedback. At first inspection, the answer seems to be that goal setting adds to feedback, but incentives do not add beyond feedback plus goal setting. This impression is based on the overall changes in productivity (measured as percentage gain of maximum possible gain), where feedback resulted in a gain in productivity over baseline of 50%; feedback plus goal setting, 75%; and feedback plus goal setting and incentives, 76%. Effect size comparisons lead to similar conclusions.

We do not believe, however, that the correct interpretation is that goal setting adds to feedback and incentives do not add further. Inspection of Figure 2 shows that across the five units, there was a large change when feedback was instituted. The early months of feedback show a very strong improvement in productivity. This improvement then begins to slow down, and finally the curve flattens out during the last months of the incentive treatment. This pattern is also clearly present within each of the five units. The negatively accelerated curve from the start of feedback to the end of incentives looks very much like a learning curve. This suggests that what may have been happening was that the units were continuing to learn how to improve productivity solely on the basis of the feedback. The possibility exists that the same amount of increase over time would have occurred if only feedback had been used. One bit of supporting evidence for this interpretation is that there was no jump in productivity at the institution of either goal setting or incentives. One would have expected a noticeable increase in the month these systems were started if they were indeed adding much to productivity.

A second and very different interpretation is that a ceiling effect was operating. One could argue that because the units were near their maximum possible level of overall effectiveness by the second month of goal setting, further increases were not really possible. Thus, both goal-setting effects and incentive effects would have been stronger if there had been more room to improve. Although this interpretation is possible, a counter argument would be that if productivity can be increased to its maximum with only feedback or feedback plus goal setting, incentives are simply not necessary. It is possible, however, that goal setting and incentives would be necessary to sustain the high levels of productivity over time.

There are other possible explanations for the lack of an incentive effect. One is that the time incentive may not have been attractive to unit personnel. Another is that they did not choose to pursue the incentives because getting time off meant they would have to make up for the lost time. Although these explanations are possible, the anecdotal data available make them doubtful. All personnel we talked to agreed that time off was indeed attractive, and there was no indication of not trying.
for time off because it would add work. The unit personnel seemed genuinely pleased when they found they had earned the time off.

There is no conclusive way of deciding between these different interpretations of the data. The only way to make such conclusions would be to replicate the study using a much longer period for each treatment to operate. Our own best estimate of the most correct interpretation is that goal setting added somewhat to feedback, but not a great deal, and that incentives did not add further. This conclusion is based on the learning-curve shape of the productivity data, the lack of a jump in productivity at the institution of goal setting and incentives, and the subjective impressions of the unit members.

It is important to discriminate in this discussion between the effects of goal setting done informally and the effects of the public, formal goal-setting program. Adding feedback could easily have led to informal goal setting, and probably did. In this sense, positive effects of goal setting are included in the feedback effect. When we speak of comparing the feedback effects with the goal-setting effects, we are comparing the effects of the public, formal goal-setting intervention with the feedback intervention.

The next research question concerned the effects of the interventions on work attitudes. The results indicate that work attitudes either improved or stayed the same across a series of attitude dimensions.

The final research question was whether the treatment effects would continue after the departure of the research team. The answer to this question is that when the system was in operation, treatment effects did continue. Although the system was being run by unit personnel, it continued to maintain productivity at essentially the highest level that it reached during the interventions. This is based on 10 months of data for Comm/Nav and 5 months for MS&D.

However, only Comm/Nav has continued with the system. In MS&D, decisions by a new manager of the supply function resulted in the system deteriorating and eventually stopping. This new manager was enthusiastic about the system when it was first presented to him, and he asked if the system could be developed for other units under his control. Unfortunately, we did not have the resources to do this. However, he was opposed to the use of the incentives, especially when used for some units under his command but not others. He did not formally stop the program, but he also did not take the steps necessary to make calculating the feedback reports a priority in the unit (outside of MS&D) that had the computers and expertise.

Group-Based Interventions

In the introduction, we argued that most of the research in industrial and organizational psychology has focused on individual performance, using simple jobs that are not representative of the majority of jobs in most organizations. Because of the complexity and interdependence of these jobs, we argued for developing quality measures of productivity and for group-based interventions. This research focused on these issues.

The jobs used in this research were indeed the more typical, complex jobs. Individuals did many different things, different individuals did different things, individuals needed to work interdependently, and measures were not readily available.

It is clear that the combination of group interventions using productivity as the criterion and feedback source were highly successful with these complex jobs. In fact, the effects were much stronger than those typically found in studies using individuals.

We also argued in the introduction that we needed to know more about designing group interventions and measuring productivity. In the process of doing the research, we learned several things pertinent to these issues.

One issue that became apparent was the importance of having the personnel who are going to be using the system be heavily involved in its development. This applies to both the productivity measurement system as well as the feedback, goal-setting, and incentive interventions. The perceptions of unit personnel were that some programs imposed from above had been ineffective. They felt that some of these programs were not designed with an appreciation for their unique needs and environment. They did the administrative work necessary to operate these programs, but otherwise the programs were sometimes ignored. It is much more effective to have heavy involvement from the unit personnel so that the final product will fit their needs, and unit personnel will not feel that it is another project imposed on them from above.

It also seems more effective to develop such programs from the bottom of the organization up. The lower levels of supervision know the most about how the unit functions; they are aware of the really critical issues, what are reasonable goals, and what are attractive incentives. In addition, these are the people who will actually make the system work. It is important to have their involvement and knowledge from the start.

It is also important to have higher level management approve the system. This should be done as the measurement, feedback, and other systems are being developed; it should not be saved until each system is complete. For example, the technique of getting approval on products and indicators before starting contingency development proved very valuable. It not only helped clarify policy earlier in the process but also helped prepare everyone for the eventual implementation. This approach of getting approval during development gave all levels of the organization a chance to learn about the system as it was being developed so that they would know how to use it when it was finished.

Another thing we learned is the importance of multiple-meeting structure, which allows for an iterative approach to the development of the interventions. Our strategy was to summarize the results of each meeting and present it at the next meeting. This was continued until consensus was reached. In this way, the unit personnel had time to think about the issues, discuss them among themselves, and be prepared to approach the question with fresh perspectives at the next meeting. Personnel needed time to adjust to the idea of completely capturing what they do in a productivity measurement system, how to structure a feedback report, how many levels of incentives to use, and so on. They needed to think about the issues, consider what implications they had, and be able to discuss the issues among themselves thoroughly.

It also proved very effective for the unit personnel who developed the measurement system to defend it when it was presented to higher management. They were knowledgeable about the subtleties of unit operation and could address questions and concerns better than the researchers. Also, it gave them a chance
to present their perceptions of optimal policy. Finally, their sense of ownership of the system was increased by their defense.

Care must be used during development so that the resulting indicators are things under the control of the unit (e.g., Hurst, 1980; Muckler, 1982). The researchers frequently needed to remind unit personnel to assess whether they had control over a given indicator. For example, at one point, Comm/Nav considered a measure that included work done by them, but also included work done by supply and flightline maintenance. They did not have complete control over such a measure, so including it in the system would decrease its motivational impact.

It is also important that the researchers develop good rapport with the operational personnel. Some personnel were initially suspicious of our intentions and had questions about our credibility. Their experience had been that outsiders came in who were not experts on their operations. These outsiders sometimes imposed programs that at best resulted in extra work and at worst had a negative impact on their effectiveness. Taking the time to really learn what they did and getting to know them greatly helped to decrease these barriers.

Earlier we discussed the importance of higher management support in the development and approval of a program; it is also important that these programs have a critical mass of higher management support during their operation to ensure continuation. One application of this principle is that it is very important that senior management be kept informed about the results of the program and fully understand it.

We did not anticipate the difficulty of keeping higher management support of the project when there was turnover in these management positions. There was generally an initial resistance to the program by the new manager. In retrospect, this is not surprising. The program was not developed by the new manager, and he or she did not have a sense of ownership of it. The implication of this point is that it is very important to take considerable time with the new manager so that he or she can be convinced that the program is indeed a sound one.

Another issue especially associated with the feedback was that management in some cases seemed to focus much more on small amounts of negative productivity information rather than on the much larger amount of positive information. For example, in some meetings where the feedback reports were discussed, a manager would focus on the few areas where the unit decreased in effectiveness and largely ignored the majority of areas where productivity had improved or was quite high. This had the effect of making those meetings a somewhat punishing experience for unit members rather than a positive experience. It is important that management be trained to conduct these meetings, both so that they will give recognition when it is due and so that they will appropriately weight positive and negative information.

There were also several issues associated with goals and incentives. We have already discussed the issue of "reportable" and "nonreportable" goals, the importance of participation in the design of the system, and the importance of joint supervision and incumbent participation in setting the goal level. For incentives, we have discussed the importance of the number of incentive levels and equity across units in what it takes to obtain the incentives.

Another issue regarding incentives is that one should expect resistance to the use of incentives, especially from supervisors. In this study there was considerable resistance. The major point advanced by those against incentives was that personnel should not get something for doing what they were already supposed to do. There was also indirect evidence that some supervisors felt such an incentive system would undermine their power and prerogatives to reward individuals and units informally. In contrast, the incumbents were much more positive about the incentive system than were the supervisors. They wanted the time off, but what was more important, they wanted some tangible form of recognition for their high productivity. There were, of course, some supervisors who were in favor of incentives and some incumbents who were not.

A final issue is that care should be taken with an incentive system so that management does not increase the level of performance needed to obtain an incentive once productivity has increased. This is a problem because it could easily lead to resentment and a sense of inequity for unit personnel. There was an attempt to do this in one unit, and the research team managed to convince the manager that this could have negative consequences. Future investigators should be prepared for this possibility.

Future Research

A number of important questions are raised by this research. One is whether the ProMES would work as well in other kinds of jobs and organizations. Because this effort was done in a military setting, it is important to know whether the approach will be as successful in a civilian setting. Although we would expect this to be the case, empirical tests are needed. One such test is currently being conducted in a civilian white-collar setting. Another issue is the generalizability of the results to units with different structures. In this study, the units had fairly specialized functions with a clear mission that was stable over time. It will be important to evaluate the system in organizations in which there is more ambiguity of mission and more rapid changes in the environment that are due to factors like rapid technological change, changes in market conditions, possibility of takeover, and so on. In addition, trying the system with organizations with structures different from the classic hierarchical structure, such as matrix organizations or decentralized flat organizations, would be valuable. We would be interested in opportunities to try the system in a variety of different organizations.

Other meaningful questions include whether the productivity measurement system would work on a larger number of units with more hierarchical levels. It is also unclear how much participation is actually necessary in the development of the productivity measurement system for the feedback to be effective. In this research the participation was considerable, but is such extensive participation necessary? It is also not clear what happens in the process of developing the system, or what is changed to increase productivity. We need to explore the changes in attitudes and behavior that occur during the development of the system and the changes that occur to produce the increases in productivity.

Finally, the incremental effects of group feedback, goal setting, and incentives must be explored with longer interventions than were used here. Only with such a design can the differential effectiveness of the three interventions be established.
In conclusion, this project has shown that productivity in units with complex jobs can be measured with the ProMES. The productivity measurement system was very successful in this application and has many advantages that make it attractive for other applications as well. The use of group-level feedback resulting from the productivity measurement system produced large increases in productivity and may well have accounted for the majority of the overall productivity gain. Goal setting seemed to add to productivity, but incentives did not add beyond feedback and goal setting. The interventions of group feedback and group goal setting clearly have an important role in productivity enhancement.

References


Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.


Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.


Received May 26, 1987
Revision received November 16, 1987
Accepted November 19, 1987