Performance management in a field service department : design and transportation of a productivity measurement and enhancement system (ProMES)

PERFORMANCE MANAGEMENT IN A FIELD SERVICE DEPARTMENT

Design and Transportation of a Productivity Measurement and Enhancement System (ProMES)

P.A.M. Kleingeld

PERFORMANCE MANAGEMENT IN A FIELD SERVICE DEPARTMENT

Design and Transportation of a Productivity Measurement and Enhancement System (ProMES)

PERFORMANCE MANAGEMENT IN A FIELD SERVICE DEPARTMENT

Design and Transportation of a Productivity Measurement and Enhancement System (ProMES)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr. J.H. van Lint, voor een commissie aangewezen door het College van Dekanen in het openbaar te verdedigen op

dinsdag 13 september 1994 om 16.00 uur

door

Paulus Adrianus Matthias Kleingeld

Geboren te Utrecht

Dit proefschrift is goedgekeurd door de promotoren

prof.dr. J.A. Algera
prof.dr. P.M. Bagchus

en door de co-promotor
dr. H.F.J.M. van Tuijl

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Kleingeld, Paulus Adrianus Matthias

Performance management in a field service department : design and transportation of a productivity measurement and enhancement system (ProMES) / Paulus Adrianus Matthias Kleingeld. - Eindhoven : Technische Universiteit Eindhoven
Thesis Eindhoven. - With ref. - With summary in Dutch.
ISBN 90-386-0433-5
Subject headings: performance management / productivity.

Druk: ICG Printing, Dordrecht

Acknowledgements

This thesis is the final product of a four-year research project, conducted in the department of Technology and Work of the Graduate School of Industrial Engineering and Management Science of Eindhoven University of Technology. The project involved the design and implementation of a ProMES performance management system in a field service department. Many people have provided support in various stages of the project. I would like to thank them all and I realize that the following list is incomplete.

I am much indebted to prof.dr. Jen A. Algera and dr. Harrie van Tuijl for providing me with challenging goals and valuable feedback in the course of the research project. Harrie van Tuijl provided a lot of additional practical support by acting as co-facilitator in design meetings, feedback meetings, and training sessions and by setting up part of the behavior modeling training program. Prof.dr. Paul Bagchus and prof.dr. Peter Sander courageously worked through the rather voluminous manuscript and provided useful comments. The graduate work of René Coolen (who acted as co-facilitator in the final stages of the design process) and Barbara Lamberts (who conducted a financial analysis in support of the design process) is much appreciated, as is the statistical advice of dr. Albert Brinkman. My colleagues at the department of Technology and Work provided an amicable and stimulating atmosphere.

I consider it an honor to have been part of the international research collaboration on the ProMES method and I very much enjoyed the discussions with prof.dr. Robert D. Pritchard, prof.dr. Uwe Kleinbeck, dr. Klaus-Helmut Schmidt, prof.dr. Henk Thierry, and their colleagues.

Not only did Nashuatec provide financial support for conducting the research, I also encountered an unconstrained and open atmosphere. I would like to express my appreciation to Harry van Gameren, M.Sc., Arjaan Berkelaar, and the other members of the steering committee (Kees Karelse, Johan van Kessel, Han Kandelaars and Hildo Segaar) for our long-running and productive cooperation. A large number of technicians and supervisors showed both effort and persistence in designing and improving the ProMES system; among them, Ron Näring, Albert Sloot, and Gert Blad deserve special recognition for their efforts as members of the design teams. The efforts of Ton van Rijswijk, who developed a computer program for providing graphical feedback, are also highly valued. Michel van Eijk did an excellent job of incorporating the complex ProMES design into the company's information system.

Finally, I wish to thank my family and friends for their interest and support.

Ad Kleingeld
May, 1994

To Ad Senior and Cocky

Contents

Chapter 1
Introduction

This thesis represents research on motivating employees through measuring task performance, providing feedback and setting goals; these are key elements of 'performance management'. Central to this research project is 'ProMES (Productivity Measurement and Enhancement System)', a method for designing performance management systems. The two main research topics relate to the generalizability of the ProMES method and the role of participation in the design of a ProMES system.

This chapter provides a brief introduction into performance management and –through an example–into the ProMES method. Furthermore, the 'design cycle' is presented as a framework for the research project. A short overview of the thesis concludes this chapter.

1.1. HUMAN RESOURCE MANAGEMENT

Organizational productivity is a function of the way three variables are managed, namely technology, capital, and human resources. The importance of an organization's human resources for achieving organizational objectives is increasingly recognized. Human resource activities, such as recruitment and selection, appraisal, compensation, and training and development, have a major impact on individual and team performance and, consequently, on organizational productivity. This is shown in Figure 1.1 within the so-called 'Human Resource Cycle' (based on Fombrun, Tichy & Devanna (1984)). It shows human performance as a function of the effectiveness of four processes (Cascio, 1989; Algera, Janssen & Van Tuijl, 1992; Tichy, Fombrun & Devanna, 1982).

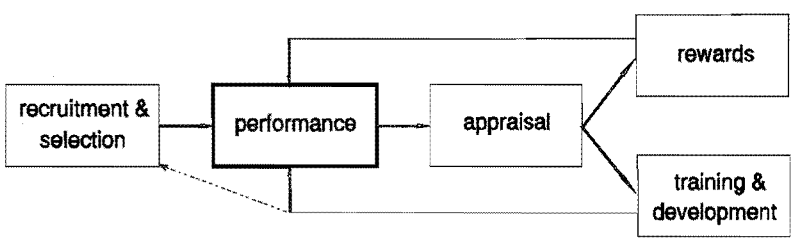


Figure 1.1. The Human Resource Cycle (Fombrun, Tichy & Devanna, 1984).

Recruitment and selection entails the definition of the organization's needs for particular positions and assessing the external and internally available pool of people to determine the best fit (i.e. choosing the people who possess qualities with which they can contribute effectively to the organization). *Performance appraisal* involves a systematic description of an employee's strengths and weaknesses, with the purpose of improving the job performance of the employee and of providing information to employees and managers for use in making decisions (setting performance goals, allocation of rewards, promotion, admission into a training program). Through rewarding competency and job performance, *compensation* stimulates that the employees who are capable of delivering the intended contribution want to join the organization, remain member of the organization, and proceed to deliver the intended contribution. *Training and development* activities are aimed at preserving and enhancing employees' competency in their jobs through improving their knowledge and skills.

1.2. PERFORMANCE MANAGEMENT

The Human Resource Cycle represents a long-term approach to controlling the competencies and performance of the organization's human resources. For example, performance appraisal and compensation processes usually operate on an annual basis. In contrast to these long-term activities, performance management systems take a short-term perspective to controlling employees' performance: their aim is to assist individuals, groups, and departments in controlling their day-to-day work through goal setting and feedback procedures (control loops for self-regulation). To illustrate the importance of performance measurement, goal setting and feedback in producing high performance, a 'High Performance Cycle' (see Figure 1.2, adapted from Locke & Latham, 1990) is discussed briefly, as are some characteristics of performance management systems which, in practice, produce a *low* performance cycle rather than a *high* performance cycle.

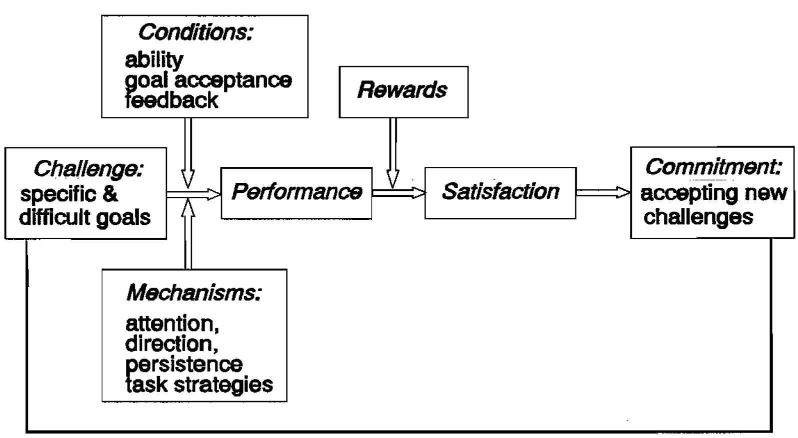


Figure 1.2. The High Performance Cycle
(adapted from Locke & Latham, 1990).

The High Performance Cycle starts with task requirements or goals that are posed to the individual. These goals should be specific and challenging. The specificity of the goals influences the direction of the behavior (what is done); the difficulty of the goal influences the effort and persistence with which the task is done. Also, goals may cause the development of new strategies for performing the task. These goals will only lead to high performance if a number of conditions are fulfilled. The individual should have sufficient ability, and he or she should accept the goals. Also, specific feedback should be provided.

Rewards are important because they provide the individual with what he or she wants or considers beneficial. These rewards can be internal (feelings of competence) or administered externally (pay, recognition). High satisfaction may be caused by having meaningful work or by rewards following high performance. In case rewards are linked to performance, high satisfaction is regarded as a consequence of high performance, rather than as a precursor to performance. Satisfaction does not influence performance directly; it only leads to high performance if it leads to increased commitment to the organization and if this commitment is to specific, challenging goals, and if these goals are associated with the necessary mechanisms and conditions (e.g. Henne & Locke, 1985).

Performance is a central element in both the human resource cycle and the high performance cycle. Therefore, accurate measurement of the performance of individuals, groups, or departments (i.e. their contribution to attainment of the organizational objectives) is essential for both long-term and short-term control of employee performance. Amongst others, this means that performance management should be approached systematically, i.e. goals and priorities at lower organizational levels should be consistently linked to the objectives of the organization as a whole.

In practice, a lot of performance management initiatives have not been very successful due to neglecting the concepts in high-performance cycle. Most problems with these initiatives fall into three categories (Algera & Van Tuijl, 1990). The first category includes sub-optimal use of goal setting. For example, goals that are set reflect a do-your-best situation, in which it is not clear what should be done to improve performance. The second category involves inadequate feedback. Although a lot of performance data is collected, very little systematic feedback is provided to the right persons, groups or departments. Thirdly, incomplete measurement often causes discrepancies between the goals of management (e.g. improving quality and maintaining quantity) and the information fed back (e.g. only information on quantity, which is often easier to collect). This causes unintended consequences (e.g. improved quantity and decreased quality) and a suboptimal overall performance.

1.3. THE PROMES RESEARCH EFFORT

A recent development in the area of performance management is the 'Productivity Measurement and Enhancement System: ProMES'. ProMES is a method for developing group performance management systems. The method was developed by Dr. Robert D. Pritchard and his colleagues at the University of Houston and Texas A&M University

(Pritchard, Jones, Roth, Stuebing, Ekeberg, 1986, 1987, 1988, 1989; Pritchard, 1990). ProMES is characterized by participative development of a measurement system, periodical feedback, and goal setting. The first field test of the system at a US Air Force base (Pritchard et al., 1989) was very successful: development and implementation of ProMES performance management systems in maintenance and warehouse units resulted in productivity increases that were much larger than those found in the literature on feedback and goal setting systems (Guzzo, Jette & Katzell, 1985).

The research project this thesis reports on is one of the first in a series, conducted by the Technology and Work/Personnel Management Group of the Department of Industrial Engineering and Management Science at Eindhoven University of Technology. This series of projects attempts to test the ProMES approach to measuring and improving productivity in a wide variety of settings. The settings differ regarding the type of 'industry' (e.g. manufacturing, health care, retail trade, education, etc.), characteristics of the units (task complexity and uncertainty, group or individual tasks, organization level, etc.) and the objectives of the system (productivity improvement, performance appraisal, etc.).

1.4. PROMES: AN EXAMPLE

In this section, we will describe the ProMES procedure for developing a performance management system. A hypothetical example, based on Pritchard et al. (1989) and Pritchard (1990), will be used to elucidate the main steps of the method. Some results from the first large-scale field test of the system by Pritchard and his colleagues (Pritchard et al., 1986, 1987, 1988, 1989) are discussed briefly. Finally, some research questions Pritchard formulated after the successful first field test of the method are presented. These questions–some of which are addressed in this thesis–represent the main areas of interest in research on the ProMES method. The theoretical background of the system and criteria for effective development and implementation of the system will be discussed in a more comprehensive way in Chapter 2.

1.4.1. Steps in the development of ProMES

Developing a ProMES system involves four steps: (1) identifying products, (2) developing indicators, (3) establishing contingencies, and (4) designing the feedback report. These steps are completed by a *design team* consisting of representatives of the unit, the unit supervisor and a facilitator. After the second and third step, a *review and approval meeting* is organized to obtain formal management approval of the system elements that have been completed.

To present these four steps concretely, we shall use a hypothetical example of a maintenance unit that diagnoses and repairs electronic equipment. The unit is responsible for repairing various malfunctioning items as quickly and accurately as possible. Periodically, the unit is inspected by a quality control unit, which determines whether it is accurately following the repair procedures delineated in the repair manual. The maintenance unit is also responsible for conducting on-the-job training. A unit member is only allowed to repair a piece of equipment if he or she has obtained a training certificate for that piece of equipment.

Step 1: Identifying products

Products are the unit's major contributions to the organization. Products can be regarded as the set of activities or objectives the unit is expected to accomplish. Products should be clearly stated: if the unit exactly does as the product says, the organization would benefit. Secondly, the set of products should be complete. Assume the design team develops the following products:

1. *Quality of repair*: the degree to which repaired items function properly after repair.
2. *Meeting demand for repairing items*: the degree to which the items are repaired quickly.
3. *Meeting training needs*: the degree to which on-the-job training needs are met.

Step 2: Developing indicators

Indicators are concrete measures of how well the unit is generating the products. There are a number of criteria for good indicators: they should cover all products and cover each product completely; they should be valid, controllable, cost-effective to collect and understandable and meaningful to the personnel in the unit. Assume the following indicators are developed:

Product 1. Quality of repair.

Indicator A: *Return rate*: percentage of items repaired that were returned, because they malfunctioned immediately after installation.

Indicator B: *Percentage of quality control inspections passed.*

Product 2. Meeting demand for repairing items.

Indicator: *Number of items repaired, divided by total number of items brought in for repair.*

Product 3. Meeting training needs.

Indicator: *Number of people qualified to work on each type of item, divided by the number of people needed to be qualified.*

Once the list of products and indicators has been completed to the satisfaction of the design team, the next step is to obtain formal approval of the list of products and indicators from management.

Step 3: Establishing contingencies

After management approval is obtained on the list of products and indicators, contingencies are established. A contingency is the relationship between the amount of the indicator and the effectiveness of that amount (see Figure 1.3 for an example). The horizontal axis of the contingency represents the range of indicator values, from the worst feasible level to the best level that is realistically possible. The vertical axis represents the range of effectiveness scores from a maximum effectiveness level of +100 to a minimum effectiveness level of −100. The contingency function shows the relationship between the indicator value obtained and the effectiveness of that value.

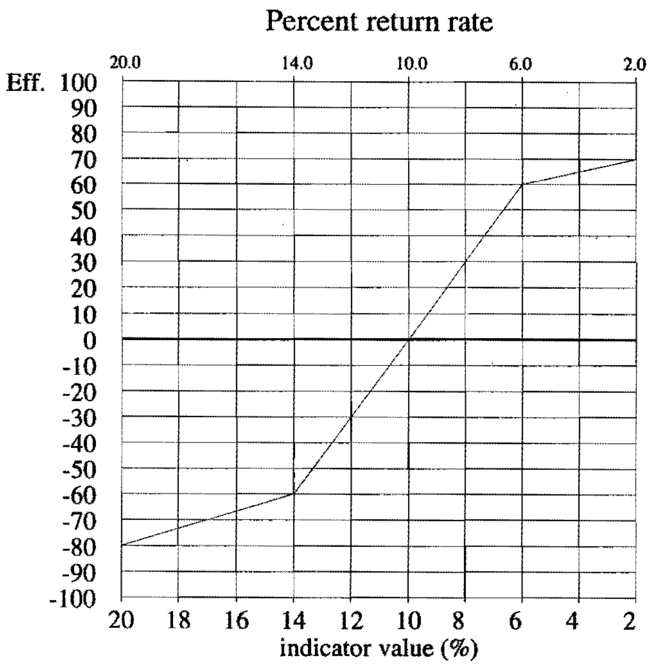


Figure 1.3. Contingency for 'Percent return rate'.

There are three steps in the development of the contingencies:

1. Establishing the maximum, minimum and neutral indicator values
2. Establishing maximum and minimum effectiveness scores
3. Drawing the complete contingency

In explaining these steps, we will use Figure 1.3, representing the contingency for 'Percent return calls', as an example.

Establishing the maximum, minimum and expected indicator values. Assume that the unit members indicate that the best return rate possible is 2% (e.g. because an average 2% of the items will fail immediately when put in use, although they have been repaired and installed properly). Let us also assume that they consider a return rate of 20% the worst feasible indicator value. Next, the neutral indicator value (or: zero point) is established; this is the indicator value which is neither good nor bad. By definition, this indicator value corresponds with a zero effectiveness score. If the neutral point was identified as a return rate of 10%, it would be so indicated in the contingency of Figure 1.3. Illustrative maxima, zero points and minima for the four indicators in the example are given in Table 1.1.

Table 1.1. Indicator value: maxima, zero points, and minima.

Indicator values	minimum	zero point	maximum
Percent return rate	20%	10%	2%
Percent quality inspections passed	80%	100%	100%
Percent repair demand met	50%	80%	100%
Percent qualified/needed	70%	100%	130%

Establishing maximum and minimum effectiveness scores. First, the maxima for each of the (in this example, four) indicators are listed. The unit members and supervisors in the design team then rank order these maxima in terms of the contribution of each maximum to the overall effectiveness of the unit. The maximum with the highest importance rank is then given an effectiveness value of +100. The other maxima are rated relative to this. For example, if the maximum of a given indicator were half as important to the effectiveness of the unit as the most important maximum, it would be given a score of +50. In an analogous way, the minima are determined. However, the most important minimum is not constrained to a score of −100. Rather, it is given the value that the group thinks is appropriate. The minimum and maximum effectiveness scores for the four indicators may look like those in Table 1.2.

Table 1.2. Minimum and maximum effectiveness scores.

Effectiveness scores	minimum	maximum
Percent return rate	−80	+70
Percent quality inspections passed	−40	0
Percent repair demand met	−85	+100
Percent qualified/needed	−60	+10

Drawing the complete contingency. After the zero points have been identified and the effectiveness scores for the maximum and minimum indicator values have been established, the complete contingency can be drawn, using these three known points as benchmarks. The contingency in Figure 1.3 for the 'Percent return rate'-indicator shows that a maximum effectiveness of +70 is reached at the best possible return rate of 2%. To be at the worst possible return rate (20%) would correspond with a minimum effectiveness score of −80. The contingency also shows that exceeding the neutral point would result in positive effectiveness. This increase is not linear: once a return rate of 6% is attained, further increases do not represent as large an increase in effectiveness. Likewise, at the negative end of the contingency, once the return rate becomes as bad as 14%, the unit is doing very badly and further decreases are proportionally not as bad.

After the design team has established contingencies for all indicators, management approval is again sought. An important issue during the second review and approval meeting is whether the contingencies accurately reflect management policy with regard to the unit. The complete set of contingencies for the hypothetical example is shown in Figure 1.4.

When the contingencies are done, all indicator values that could reasonably occur can be converted into effectiveness scores, which enables *direct comparison of the* (sometimes completely different) *indicators*. Moreover, the *relative importance of the indicators* is reflected in the system by the steepness of the slope of the contingency, which is determined by the choice of maximum and minimum effectiveness. A steep slope implies that variations in the indicator values attained result in large variations in effectiveness; a less steep slope implies that variations in the indicator values result in smaller variations in effectiveness.

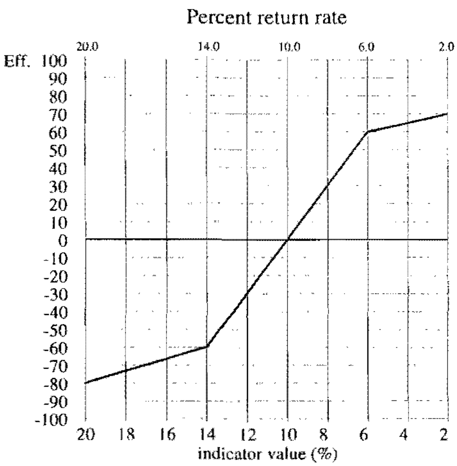
The second indicator for the 'Quality of repair'-product is 'Percent quality inspections passed'. The contingency for this indicator has an expected level of 100%, indicating that carrying out repairs in less than full accordance with the repair manual results in negative effectiveness.

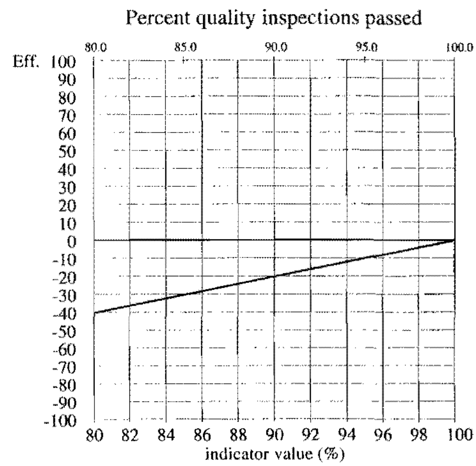
The indicator for the second product ('Meeting repair demand') is 'Percent demand met' (the number of units repaired divided by the number requiring repair, expressed as a percentage). The contingency shows that indicator values between 70 and 90 percent represent hardly different effectiveness levels, but that slipping below 70% results in increasingly negative effectiveness, whereas attaining indicator values above 90% results in large increases in effectiveness.

The indicator for 'Meeting training needs' is 'Percent qualified/needed' (the number of unit members qualified to repair equipment divided by the number needed). It is possible to attain indicator values above 100% by having more members qualified than needed. The contingency for this indicator becomes flat at 110%, hereby indicating that having more than 110% is no more effective than having 110%. So, once some flexibility is created by having 10 percent more qualified personnel than needed, having additional trained personnel is not important.

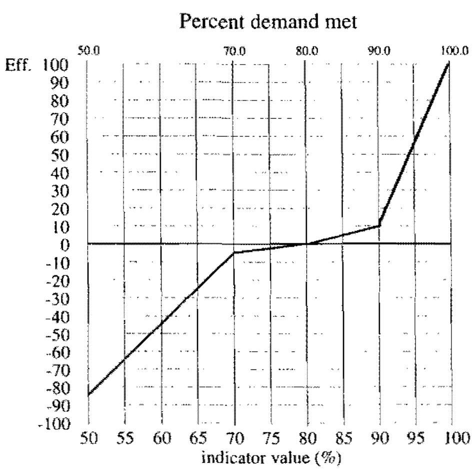
These contingencies illustrate that contingencies are typically *non-linear*. These non-linearities take into account such phenomena as 'diminishing returns'. Once a unit achieves a high level of productivity on one aspect of the work, it is frequently better for the organization to work on improving something that the unit is not doing as well, rather than to continue to improve something that is already at a high level. For example, if the maintenance unit was operating with a good return rate (e.g. at 6%/+60), it might be better to try to improve meeting training needs (e.g. at 80%/−40) rather than attempt to further improve its return rate.

Quality of repair





Meeting repair demand / training needs



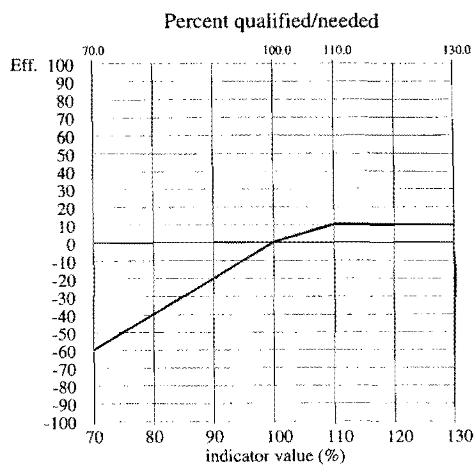


Figure 1.4. Set of contingencies for the maintenance unit.

Step 4: Designing the feedback report

The final step in developing the ProMES system involves designing a feedback report which is a formal description of the unit's productivity for the past period. The feedback report is the basis for discussions about improving productivity during periodical feedback meetings. Decisions have to be made regarding the frequency of the feedback (depending on the job cycle and the feasibility of preparing the feedback reports), the timing of the feedback (as soon as possible after the end of the reporting period), which persons in the company will receive the feedback report (at least unit members and levels of management that have been involved in the development of the system), and the extent to which the feedback is public or private (in general, individual feedback should be given privately; group feedback might be public).

The basic information included in the feedback report would be: the list of products and indicators and the level of each indicator for the period along with its associated effectiveness scores, the total effectiveness score for each product (which is the sum of the effectiveness scores of the products' indicators), and the overall effectiveness score (which is the sum of the effectiveness scores of the individual indicators/products). The feedback report of the maintenance unit is presented in Figure 1.5.

PRODUCTIVITY: MAINTENANCE UNIT
Period: January 19xx

	Indicator value		Effectiveness score
1. Quality of repair			
A. Percent return rate	6%		+60
B. Percent quality control inspections passed	95%		−10
Total effectiveness: Quality of repair		+50	
2. Meeting repair demand			
A. Percent demand met	90%		+10
3. Meeting training needs			
A. Percent qualified/needed	80%		−40
Overall effectiveness			+20

Figure 1.5. Feedback report of the maintenance unit.

In addition to the feedback report shown in Figure 1.5, other types of information can be provided, such as:

- historical data: graphs, providing historical information on productivity over a long period of time (e.g. since the start of the program) to spot trends;

- comparative data: data comparing the previous period with the current period and showing the amount of change;
- priority data: a listing of effectiveness increases derived from projected improvements on each of the indicators.

The *ability to simply sum effectiveness scores* is one of the major advantages of the system. Because the contingencies already reflect the relative importance and non-linearity of the indicators, a simple summing reflects the overall effectiveness of the unit. In case feedback reports are available for several units, it is possible to *compare the productivity of* these *units*, even if they perform completely different functions. An *aggregate* effectiveness score can be calculated for the productivity of all units together.

Implementation of the system: feedback meetings

At the end of each measurement period (e.g. a week or a month), the indicator data are collected and a feedback report is compiled. During a feedback meeting, the feedback report is discussed by all unit members (or at least a group of representatives from the unit) and the unit supervisor. The feedback report is reviewed and areas where productivity has increased or decreased are explored and the causes of these changes are identified. Strategies for maintaining or improving overall effectiveness are discussed and the effects of the implementation of these strategies are monitored.

Needed organizational resources

The main resource needed for developing and implementing ProMES is time of unit personnel, facilitators, supervisors, and (to a lesser extent) management. The design team develops the system in a number of meetings, which are typically between one and a half and two hours long and which are held every two weeks (variations from this schedule are possible if required by the situation). Pritchard (1990, p. 69) provides a rough estimate of the number of meetings needed to develop a ProMES system if the facilitator has some experience (see Table 1.3). The actual time needed will vary, depending on a number of factors, such as the experience of the facilitator(s), the complexity of the unit's work, the degree of cooperation within the design team, and the number of measures already in use.

Implementation of the system will require collection of the indicator data and preparation of the feedback reports. Periodically, a meeting must be held between unit personnel and the supervisor to discuss the feedback report covering the past period.

Other interventions with ProMES

Essentially, the ProMES system consists of two parts, a measurement system and a formal feedback system. In addition to feedback, the ProMES system facilitates the use of other motivational techniques, such as formal goal setting and incentive programs.

For example, the availability of a single index of productivity is very useful for setting specific, quantitative goals. According to Pritchard (1990), the appropriate order of these interventions would be first to develop the measurement system and then institute the formal feedback system. This feedback system should operate for several months, so that necessary adjustments to the system can be made. After this period, goal setting could be added, but only if additional increases are expected which would be worth the additional effort to develop the goal setting system. After that, incentives could be added if additional positive effects are expected.

Table 1.3. Estimated number of design meetings.

Development step	number of meetings
Review of the program	1
Identifying of products	2
Developing and refining indicators	5-8
Management approval (products & indicators)	1
Developing contingencies	3-4
Final management approvement	1
Designing the feedback report	2
Overall number (1½-2 hours each)	15-19
Total meeting time	26-33 hours
Total development time (2 weeks between meetings)	6-8 months

1.4.2. Research findings

The first large-scale field test of the ProMES method was undertaken at a US Air Force base in the southwest of the United States (Pritchard et al., 1986, 1987, 1988, 1989). Five organizational units (one maintenance unit and four units in the supply area) took part in the project. The mean number of personnel was 32 in the maintenance unit and 15, 15, 13, and 7 in the supply units.

All five units developed a productivity measurement system according to the ProMES method. This was followed by an eight-month baseline period during which productivity data were collected, but not fed back. Next, feedback was provided to each unit for five months. During the next five months, formal goal setting was added to the feedback. Finally, incentives (in the form of time off from work) were added to the feedback and goal setting; this final treatment also continued for five months.

Effects on productivity

Figure 1.6 shows the combined effectiveness scores of the five units during baseline and treatments. Taking baseline effectiveness as a starting-point, the percentage increase compared to the maximum possible increase (combined maximum effectiveness minus combined baseline effectiveness) shows a 50 percent increase for feedback, a 75 percent increase for (feedback plus) formal goal setting, and a 76 percent increase for (feedback plus formal goal setting plus) incentives. These results indicate a major increase in productivity. Statistical analyses show that the difference between the feedback mean and the baseline mean was significant as was the difference between feedback plus formal goal setting and feedback alone (see Pritchard et al., 1989). Adding incentives to feedback plus formal goal setting did not cause further significant improvements.

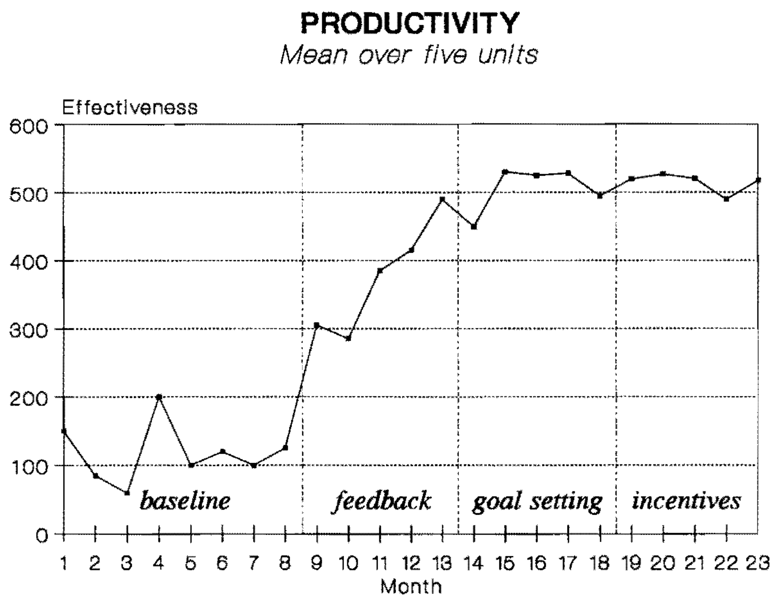


Figure 1.6. *Mean productivity of the five US Air Force units* (source: Pritchard et al., 1989).

These increases in productivity were not caused by changes in number of personnel. By the end of the project, the productivity gains that occurred were achieved with an equal number of personnel in the maintenance and a decreased number of personnel in the supply area. Also, by the end of the treatments, overtime was less than one-third of what it had been during baseline.

Control group data, collected on several units within the maintenance, supply and some other departments, generally showed no change in productivity during the treatment periods. So, the effects on productivity in the experimental units cannot be explained by wider organizational changes in productivity.

Finally, the productivity increases cannot be explained by the presence of a Hawthorne effect. This effect would entail that merely giving attention to a group (e.g. by singling it out to participate in a research project) would cause increases in productivity, irrespective of the treatments administered. In the Air Force project, special attention was given during the entire development of the system, which is a quite long and intensive process. Any productivity increase due to a Hawthorne effect should therefore have occurred during this development period and before the start of the baseline period. Consequently, the productivity increases reported here were not contaminated by a Hawthorne effect[1].

Pritchard et al. (1988) also calculated the effects of feedback, goal setting, and incentives with the procedures from the Guzzo et al. (1985) meta-analysis. This means that the *d* statistic (Cohen, 1969) was used, whereby the mean difference between conditions (e.g. feedback and baseline) is divided by the value of the pooled within-groups standard deviation. Main results are shown in Table 1.4.

Table 1.4 shows that the ProMES effect sizes are much larger than those found in Guzzo et al.'s (1985) meta-analysis (the feedback-, goal setting- and incentive-effects are 7, 6, and 8 times as large, respectively). Moreover, the mean ProMES effects are at least 3.5 times as large as the largest effect reported in any of the studies Guzzo et al. (1985) reviewed.

Table 1.4. Effect sizes of feedback, goal setting, and incentives.

	Effect	Feedback	Formal goal setting (plus feedback)	Incentives (plus feedback & goal setting)
ProMES US Air Force effect sizes	Mean	2.44	4.54	4.68
	Range	1.75 - 3.70	2.92 - 6.24	2.93 - 6.53
Guzzo, Jette & Katzell (1985) meta-analysis effect sizes	Mean	0.35	0.75	0.57
	Range	0.08 - 0.62	0.57 - 0.93	-0.10 - 1.24

Attitude data

During baseline and at the end of each of the three treatments, a questionnaire, measuring seven categories of attitude data, was administered to unit members and first line supervisors in each of the units. The overall results, shown in Table 1.5, indicate that job attitudes under the treatments were as favorable or more favorable than before the treatments. There were no significant differences between the treatments.

[1] Some evidence was found for the assumption that the development of a ProMES system may lead to productivity improvements in itself. Therefore, the baseline effectiveness values may be inflated and the effect sizes for later interventions (feedback, goal setting, and incentives) are probably conservative (Paquin, Jones & Roth, 1992).

Table 1.5. Summary of attitude results.

Attitude scale	result
Job satisfaction	increase
Morale	increase
Turnover intentions	no change
Clarity of objectives	no change
Individual role clarity	no change
Evaluation clarity	increase

Data were collected on the reactions of unit members and first line supervisors with different aspects of the system. Reactions to the system were extremely positive (65 percent of the respondents gave a positive evaluation, 28 percent a neutral one, and only 7 percent responded negatively to the system).

1.4.3. Research questions

After the successful first field test of the method, Pritchard et al. (1989) formulated a number of suggestions for future research that represent the main areas of interest in research on the ProMES method. Some of these research questions are discussed briefly.

Generalizability of the approach. The first field test of the ProMES system took place in a military setting in the United States, in a classical hierarchical organization, in units with specialized functions and a clear and stable mission, that carried out maintenance and warehouse tasks. Several questions regarding the generalizability of the method arise, such as: Does the system work as well in nonmilitary settings? Will the method fit other cultures than the Northern American one? Can the system be effective in organizations with structures different from the classical hierarchical structure? Will the method work in organizations/units confronted by rapid changes in the environment? Can successful systems be developed for other kinds of jobs (e.g. white-collar, managerial)? Developing and implementing ProMES systems in different settings would provide empirical tests of the generalizability of the approach.

Importance of participation in development process. As shown in Table 1.3, the process of developing a ProMES system is rather labor intensive and lengthy. While this seems to be beneficial for producing a high-quality system which is understood and accepted by unit members, it would be useful to know whether the development process could be shortened without decreasing the effectiveness of the system. A related issue is the 'transportability' of the system. If the system were developed by one unit and transported to another comparable unit (possibly with minor modifications), could comparable results be achieved or would the lack of participation in latter setting prevent the system from being successful?

Reliability of system development. The ProMES system attempts to identify organizational policy with regard to the unit in question. This policy is inherently subjective. An important issue is how accurately the system reflects this policy. Pritchard et al. (1989) give two examples. The reliability of system development should be assessed. Independent development of the system by separate units doing the same work could serve as a basis for assessing the agreement in products, indicators, and contingencies. The accuracy of the contingencies could be assessed by having supervisors evaluate productivity scenarios and comparing their evaluation with the ProMES results that would be obtained in each of the scenario's.

Mechanism through which the system works. It is proposed that the system affects productivity through increased motivation, and some specific mechanisms such as role clarification have been suggested. However, future research needs to identify the specific processes that occur.

Longevity of productivity gains. Although the initial productivity gains in the Air Force project were impressive, long term effects on productivity could not be monitored. It would be worthwhile to investigate how long the effects of the system last, and what conditions influence the longevity of the productivity gains.

Aggregation. In theory, the system could be developed for all units of a very large organization, and through an aggregation procedure, provide a productivity index for the entire organization. Furthermore, comparison of the productivity of the different units would be possible (for example, by comparing the percent of maximum score each unit has achieved). It would be worthwhile to try this in an organization.

This thesis addresses the first two research questions from the previous section. First, the generalizability of the approach that was very successful in the US Air Force setting is tested by applying it in a very dissimilar setting: the field service department of a Dutch supplier of office equipment, named 'Nashuatec'.

Secondly, the importance of the participative development procedure was tested. Because the field service department consisted of 14 comparable units, it was decided to develop the system participatively in two units and transport it to the remaining units, i.e. implement it without going through the development process. This reflects the second aspect of the second research question: the transportability of the system.

Providing an answer to these research questions involves, among other things, the design of a ProMES performance management system and an evaluation of the effectiveness of the design process and the design. In the next section, some methodological issues pertaining to this type of research are discussed.

1.5. CHARACTERIZATION OF THE RESEARCH PROJECT

The project this thesis reports on can be characterized as a form of applied scientific research as opposed to fundamental scientific research. These two types of research will be contrasted. Subsequently, the design cycle will be discussed as the framework for the project.

1.5.1. Fundamental versus applied research

In fundamental research, the research problem originates from theoretical knowledge. The objective of fundamental research is developing or testing a theory. The core model of fundamental research is represented by De Groot's (1961) well-known cycle of empirical research (Figure 1.7).

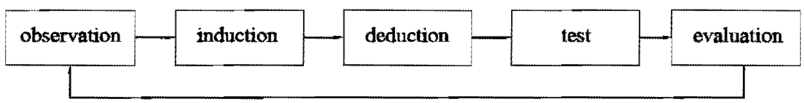


Figure 1.7. The 'empirical cycle' (De Groot, 1961).

The empirical cycle is used in solving 'knowledge problems'. 'Observation' entails the collection of empirical data. Through 'induction', hypotheses are formulated based on these data. In a process of 'deduction', verifiable predictions are made that are tested by confronting them with new empirical data. The 'evaluation' consists of an interpretation of the results of the test (are the hypotheses supported?).

In applied research, the research problem originates from a practical problem, which can range from a somewhat general societal problem to a very concrete problem within a given organization. Applied research is aimed at determining causes of the problem, identifying relevant circumstantial factors, and designing and implementing a solution. This does not mean that existing theoretical knowledge derived from the fundamental research does not play a role when carrying out applied research. In case existing knowledge–required to determine problem-causes or to design solutions–is insufficient, theory development may be necessary within the framework of applied research (Florusse & Wouters, 1991).

Raaijmakers (1993, p. 19) argues that it would be incorrect to expect applied research to contribute directly to theory-development: "If so, it would be so much gained; however, it is not aimed at doing so. Applied research is aimed at designing and evaluating solutions for tangible technological and societal problems. The value of the research should be judged by its contribution to solving those problems." In summary (p. 17): "Using applied research, one does not design a theory; using theoretical research, one does not design." In the context of industrial engineering, the most important criterion for evaluating the design of control systems (such as performance management systems) is their ability to meet the objectives specified within the organizational context (Vosselman, 1993). So, the primary contribution to be gained from an evaluation of the effectiveness of the design is an improved redesign of the system. In addition, the pragmatical knowledge gained may be useful in comparable cases in the future. Contribution to theory development will probably be limited to some ideas or hypotheses of a more exploratory nature.

1.5.2. The design cycle concept

The design cycle (Roozenburg & Eekels, 1991) is a core model in applied research. Although design methodology originates from the engineering sciences, the design cycle concept can be applied to 'products' in the broadest sense of the term (e.g. a paper clip, a bicycle, or a personnel selection procedure (e.g. Ridderbos, 1992; Roe, 1989)). The design cycle is depicted in Figure 1.8. The design cycle will be used as a framework for describing the phases in the design of a performance management system according to the ProMES method. In this section, each step is described in general terms (based on Roozenburg and Eekels, 1991).

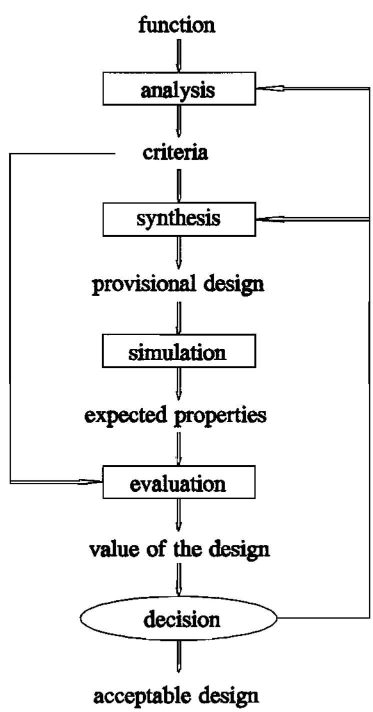


Figure 1.8. The design cycle (Roozenburg & Eekels, 1991).

Analysis. The starting point in the design process is the function (the intended behavior in the broadest sense of the term) of the product to be designed. Although it is not always possible to give a detailed specification of the function the product should fulfill, there should at least be some general notions about this function, for the designer to know what to design. In the analysis phase, the designer forms a idea of the problems existing in connection with the product to be designed. Criteria are formulated that should be met by the system, given its function. Also, the constraints posed by the context in which the product should operate, are taken into account. The criteria and constraints form the *'list of requirements'*, which should be formulated as concretely as possible.

Synthesis. In the synthesis step, a *provisional design* of the system is generated. The design is an integral solution for the problems posed in the analysis step. The term 'integral' signifies that the design is more than a simple summation of all kinds of solutions for sub-problems: it is a total solution which, as it were, eliminates all sub-problems. This step of the design process is essentially a creative activity, as the solution that is produced cannot be derived from available knowledge by deduction. This does, however, not imply that available knowledge about the field should not be used as a starting point for this creative process.

Simulation. Simulation is the judging of the behavior and properties of the designed product, using reasoning or model tests, before the actual production and use of the product. A variety of methods can be used in this step, ranging from abstract (e.g. mathematical models) to concrete (e.g. mock-ups). Yet, many simulations are based on experiential knowledge. Because products are designed to have a certain effect when they are used in a certain way, while influenced by certain contextual factors, a behavioral model of the product is preferred to simulate how the use of the design and the context in which it is used affect its functioning. Whatever model is used, the simulation should result in expectations regarding the actual properties of the product, in the form of conditional predictions.

Evaluation. In the evaluation step, the value of the provisional design is determined by comparing the expected or observed characteristics with the desired characteristics as defined in the list of requirements. Some differences will always occur, so one should judge whether these differences are acceptable or not. This is often not an easy task, since the provisional design typically fulfills some of the desired properties better than others. Usually, designing involves compromising between (partially) conflicting requirements.

Decision. Finally, a decision is made: either to continue (e.g. taking the design into production) or to try again and generate a better design. The first provisional design is usually not an optimal one. So, one has to return to the synthesis phase to improve it. This iteration may have to be carried out more than once. It may also be necessary to return to the analysis step and redefine the list of requirements. Exploring the possible solutions usually aids in gaining insight into the true nature of the design problem.

The iterative structure of the design process

Both the design and the list of requirements are developed iteratively. This iterative process is shown in Figure 1.9. When carrying out the steps described above, comparisons are continually made between the results that are accomplished so far and the desired results. The experiences gained in a cycle are fed back to both the design and the problem definition and the list of requirements.

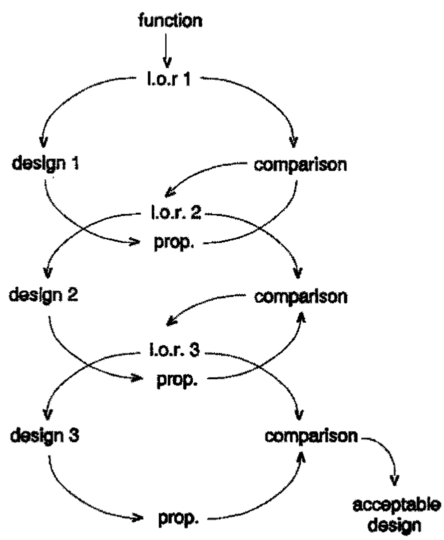


Figure 1.9. The iterative structure of the design process (Roozenburg & Eekels, 1991).
(l.o.r.= list of requirements; prop.=properties)

Based on the above discussion, the general characteristics of the research project this thesis reports on can be summarized. The research problem originates from a practical problem, which in this case exists within a specific organizational context. In analyzing the problem and designing a solution, existing empirical knowledge is used (or generated in the course of the research). Designing a solution for the problem is an iterative process. Based on criteria the design should fulfill, a provisional design is generated and its behavior is simulated. Based on this simulation, the value of the provisional design is determined and a decision is made whether the design is acceptable and final, or whether redesign activities, in which the provisional design and/or the list of requirements are adjusted, are necessary. In evaluating the final design, its contribution to solving the practical problem is of prime importance. In addition, the evaluation may produce pragmatical knowledge for use in future design processes and (in some instances) impulses toward theory development.

1.6. OVERVIEW OF THIS THESIS

Chapter 1, the present introductory chapter, provided a brief introduction into performance management and into the 'Productivity Measurement and Enhancement System (ProMES)', a method for designing performance management systems. The design cycle concept was introduced as a framework for the research project.

In Chapter 2, the theoretical and practical basis for the research project is presented. In the first part of this chapter, central concepts in the project (productivity, feedback, goal setting and participation) are discussed. In the second part of the chapter, a normative model for designing and evaluating ProMES systems is presented. This model consists of design criteria regarding goal setting and feedback procedures in general, design criteria from the ProMES method regarding the design and implementation of a ProMES system, and constraints posed by the organizational context of the system.

Chapter 3 provides an outline of the study. The research objectives, research questions and research design are discussed. The research setting–the field service department of a large Dutch supplier of photocopiers named Nashuatec–is described and the main design problems to be solved are clarified. The research conditions are described in terms of the participation concept. Finally, the design cycle framework (introduced in Chapter 1) is applied to the development of a ProMES system.

Chapter 4 represents the elaboration of the first research question of the project regarding the generalizability of the participative ProMES approach. The participative design process and the implementation of a ProMES system in two service regions of the Nashuatec organization is described in detail, with special attention to the way various design problems were addressed. The final design is presented. As for effects, the chapter mainly focuses on quantitative productivity effects.

The second research question, concerning the transportability of the system is examined in Chapter 5. The design of the transportation of the ProMES system to six service regions, which involved a feedback and goal setting training-program for the region supervisors and bilateral feedback meetings, is discussed as are the effects on productivity. A cost-benefit analysis is presented too.

Chapter 6 deals with a more qualitative evaluation of the effectiveness of the control-loop design. The first part of the chapter focuses on the extent to which the design criteria and situational constraints from the normative model have been met. Part of this evaluation was based on two questionnaires (regarding use of feedback and goal setting and regarding the effectiveness of supervisor behavior during the bilateral feedback meetings). In the second part of the chapter, two examples of interesting by-products of the ProMES program at Nashuatec (use of ProMES information in the performance appraisal of technicians and in the pay-for-performance system of the supervisors) are given.

In Chapter 7, the conclusions from this research project are summarized and some recommendations are made for future research on performance management and the ProMES method.

Chapter 2
A theoretical framework

In this chapter, the theoretical basis for the research project is presented. The central concept of productivity is defined and the use of productivity measurement as a motivational tool is discussed. Likewise, two motivation theories which are at the core of performance management and the ProMES method–goal setting theory and feedback theory–are reviewed. Main findings regarding effects of goal setting and feedback on performance and moderators are recapitulated. The concept of participation, which also plays an important part in the ProMES method, is defined and effects and mechanisms are reviewed. A normative model for designing a ProMES system–called the accepted control loop model–is presented, which is used to guide the specific interventions made within the framework of the ProMES method. Three types of design criteria and constraints are distinguished: general design criteria based on the goal setting and feedback theories, design criteria from the ProMES method, and situational constraints posed by the organizational context in which the system is to be designed and implemented.

2.1 PRODUCTIVITY

2.1.1. Introduction: productivity is important

Since the late 1970s, there has been a great concern for productivity. On a national level, productivity growth is considered an important factor in controlling inflation, conserving natural and human resources, decreasing the real costs of goods, and increasing the standard of living (e.g. Kendrick, 1984; Mali, 1978; Riggs & Felix, 1983; Tuttle, 1983). At the level of industries and individual organizations, productivity and productivity growth are also important. When the productivity growth of an industry or a firm is higher than that of its competitors, that industry or firm survives better (Sink, 1985; Tuttle, 1983). The economic integration of the European Community poses heavy pressures on the productivity of organizations, and in many areas organizations are trying to improve their competitive position (Pritchard, Algera, Janssen, Van Tuijl, 1990). Lastly, productivity and productivity growth are important to individuals. Increased productivity results in better use of time and more leisure time. It is a key to advancement in organizations (Kendrick, 1984). Productivity is also a central aspect of self-fulfillment and self-respect (Pritchard, 1992).

2.1.2. Defining productivity

While there is agreement that productivity is important and why it is important, there is very little agreement on what the term productivity means (e.g. Campbell & Campbell, 1988a, 1988b; Pritchard, 1992; Tuttle, 1983). It has been used interchangeably with such diverse concepts as output, motivation, individual performance, organizational effectiveness, market share, production, profitability, quality, net earnings per share, production cost/standard cost and so on (Pritchard, 1990; Sink, 1985).

Within a broad range of opinions about how to define productivity, the majority of authors, however, agree that the term productivity should be limited to efficiency or a combination of efficiency and effectiveness (e.g. Mahoney, 1988; Tuttle, 1983). A definition of productivity is: the ratio of the outputs generated by a system relative to the inputs needed to create those outputs (Mahoney, 1988; Sink, 1985). Effectiveness is typically defined as the ratio of output to a standard or goal, e.g. tons of steel expressed as a percentage of the goal for that month. Whereas there is agreement on this definition of effectiveness, the definition of the efficiency concept is a more complicated matter. Some authors define efficiency as the ratio of output to input, e.g. monthly number of tons of steel produced divided by number of personnel hours used to produce that output (e.g. Mahoney, 1988), which is the same as the output/input ratio used in the definition of productivity. Others refer to efficiency as the ratio of two inputs: the standard input and the actual input, e.g. the number of personnel hours calculated by the number of personnel hours actually used (e.g. In 't Veld, 1985). In 't Veld (1985) shows that this is only a seeming contradiction: (translated from Dutch) "(...) in all disciplines, effectiveness is considered the ratio of two outputs and productivity is the ratio of output and input. However, this changes when the term efficiency is used. In many cases, efficiency and productivity are used as synonyms for the same ratio, namely output/input. Those who make this distinction consider a system more efficient, when it has a higher productivity. In that case, efficiency is the ratio of two productivity numbers. However, in most cases, *two systems with the same goals* are compared: the outputs (goals) are equal and a higher productivity is the consequence of a smaller input. So, the higher productivity (output/ input) is the result of a better efficiency" (input/input, in In 't Veld's terms).

The definition of productivity used depends to a large extent on the disciplinary perspective involved. Tuttle (1981, 1983) identifies five distinct perspectives: the economist, the accountant, the industrial engineer, the psychologist, and the manager. Whereas the narrow definition (productivity equals efficiency) is dominant in for example the economist approach, the broader definition (productivity is a combination of effectiveness and efficiency) is more dominant in for example the psychologist's and manager's approach. An interesting comparison involves the industrial engineer's view and the psychologist's view. The industrial engineer views productivity as an output to input ratio. It is based on the model of a machine (or: man-machine system), where productivity is the ratio of useful work (output) divided by the energy used to produce the work (input). The psychologist, on the other hand, considers productivity in a behavioral perspective, and focuses primarily on the aspects of productivity that the individual can control (e.g. Guzzo, 1988; Ilgen & Klein, 1988). The assumption is that by changing individuals' behavior, productivity will be changed.

The only way to avoid the insoluble problem of establishing the one best way to conceptualize or measure productivity, is determining the purpose for measuring productivity, since the proper conceptualization and measurement of organizational productivity depends on the purpose intended. No one approach is best, and different purposes suggest different approaches (Pritchard, 1992). Five major categories are identified.

- Comparing large aggregations of organizations, e.g. comparing national economies such as the United States with Japan or comparing the aircraft industry with the pharmaceutical industry.
- Evaluating the overall productivity of individual organizations for comparison with each other or with some standard, e.g. an organization comparing itself to similar organizations to assess its competitive position.
- Gaining management information. Here the focus is on (a large part of) a single organization. Productivity deals with the overall functioning of the human/technological system. Such measurement is used by top management for strategic planning and policy making.
- Controlling parts of the organization. Here the focus is typically on a single function within an organization, for which specific measures unique to that function are used. Examples are production engineering, quality control, physical distribution, logistics and inventory control.
- Use as a motivational tool. The objective is to improve productivity, and the assumption is that if individuals change their behavior appropriately, their productivity will increase.

The proper conceptualization of productivity (disciplinary perspective taken, efficiency and/or effectiveness approach used), depends on the purpose of measurement. Each of these purposes requires a productivity measurement system that is different from the others.

2.1.3. Productivity measurement as a motivational tool

In Section 1.4, the ProMES method was described, in which productivity information is measured and fed back to personnel with the objective of increasing productivity. ProMES is an example of using productivity measurement as a motivational tool. It is therefore interesting to describe this use of productivity measurement in some more detail and look at some important differences between measuring productivity for use as a motivational tool and the other four purposes (Pritchard, 1992).

When using productivity measurement as a motivational tool, one assumes that organizational personnel have a substantial impact on the productivity of the organization. While the technical subsystem is also important, the focus is not on that part of the system directly, but on how the personnel uses the technical subsystem (e.g. Wilson, Neely & Aggarwal, 1993). A productivity increase would occur through changes knowledge and skills or through changes in motivation, where motivation includes the amplitude, persistence, and direction of behavior (e.g. Campbell & Pritchard, 1976). This means that personnel would exert more effort and be more persistent in these efforts, their efforts would be more directly related to organizational goals, and, if necessary, they would develop and use better work strategies.

In the context of the individual employee (or group of employees) as the unit of analysis, it is useful to distinguish among four related variables: behavior, performance, effectiveness, and productivity (Campbell & Campbell, 1988b). Behaviors are the observable things people do while at work. Performance refers to the level of expertise with which an individual executes behaviors that are relevant for one or more goals of the organization. Effectiveness refers to the outcomes of performance (e.g. as a result of solving a machine malfunction, a group member may positively influence the amount of production turned out by his or her group). Productivity refers to the ratio of effectiveness to the cost of enhancing that level of effectiveness (so, the broader definition of productivity is used).

An important difference between measuring productivity as a motivational tool and the other purposes relates to the issue of controllability (Pritchard, 1992). In the other applications the combined effects of the personnel and the technological subsystem (and, in some approaches, the environment) are assessed. Measuring productivity for motivational purposes implies measuring those aspects of the organization's productivity that the personnel can control. Consequently, effects of factors that personnel cannot control should be excluded from the measurement system. Secondly, all important aspects of the work should be measured. For example, if both quality and quantity are important, both should be measured. In the other approaches on must usually be satisfied with an incomplete set of productivity measures, since measuring all important functions is often not feasible (e.g. if one wants to compare organizations doing very different things, only those measures can be used that are common across organizations and for which data are available (Pritchard, 1992)). Thirdly, whereas in the other approaches efficiency is often the predominant measure, both efficiency and effectiveness measures should be used in the motivational approach. Meeting organizational goals without consideration for the resources used to do so may be detrimental to the organization. On the other hand, efficient use of resources to produce outputs that do not meet quality requirements is not in the organization's best interest. Substitutions among inputs and outputs are acceptable as productivity enhancement as a means of accomplishing goals (Mahoney, 1988). Although the relative importance of efficiency and effectiveness may vary according to the technology and level and type of organization (Tuttle, 1983), most organizations must be concerned with both. Pritchard (1992), therefore, defines productivity in such a way that both efficient use of inputs to produce outputs and producing outputs that meet organizational goals are included: "Productivity is how well a system uses its resources to achieve its goals".

The term 'productivity' will be used in this thesis as a collective noun for all quantitative effects of the use of the ProMES measurement and feedback system. In the light of the issues discussed above, this may seem to be a simplification, since only a minority of the measures generated by the ProMES method will be 'pure' indices of productivity (such as 'units produced per personnel hour). For example, 'units produced' would be an effectiveness index, 'time spent per product' would be an efficiency- (in the input/input sense) indicator, and 'percentage of critical safety behaviors exhibited' would be an index of performance. Continuously making this distinction would be more confusing than illuminating. Therefore, the term productivity will be used when discussing quantitative effects of ProMES programs.

Finally, the use of the term effectiveness on the vertical axis of a ProMES-contingency suggests that all indicators that are 'put into' the contingencies are effectiveness measures. This is not the case. The term effectiveness refers to the comparison of the indicator value and the 'expected' indicator value (which corresponds with the zero point in the contingency). So, technically, the scores generated by the system are effectiveness measures. As Pritchard (1990, p. 148) states: "Even if the indicator is actually an efficiency[1] measure such as units produced per personnel hour, the contingency is a way of showing how the unit's level of efficiency compares to expectations. In other words, the effectiveness of that level of efficiency".

2.1.4. Desirable characteristics of a productivity measurement system

A productivity measurement system which is used with the objective of improving productivity through behavior change should meet a number of general design criteria. The following list of criteria is based on Pritchard (1990) and Pritchard (1992).

A productivity measurement system should provide an *overall index of productivity*. A single index has motivational value, because it provides personnel with a sense of improvement or decrement. This allows them to see the results of their efforts and strengthens the tie between behavior and outcomes more clearly than a large number of incomparable measures. A single index also facilitates attempts at organizational change and the evaluation of these attempts. For example, when using a goal setting and incentive system in connection with a productivity measurement system, the single index can readily be used as the basis for setting the goal and awarding the incentives. Multiple measures, on the contrary, require multiple goals and make awarding incentives difficult.

In addition to an overall index, a productivity measurement system should use *subindices* that provide information on the separate functions of the unit measured. To be able to compare different functions, the subindices should be *on a common metric*. Personnel can then see how they are doing on the different functions and direct their behavior accordingly. Information on subindices is also useful for identifying problem areas, determining priorities, and devising strategies to increase productivity.

The system should be *valid* in four respects. First, the system should be *complete*. All important aspects of the productivity of the organizational unit should be measured, not just those for which measures are readily available. Secondly, 'the right things' should be measured. The system should correctly *reflect organizational policy* for the unit in question. This relates to the third aspect of validity. Not all functions of a unit are of equal importance to the organization. These differences in importance must be preserved in the measurement system by some method of ***importance weighting***. The final aspect of validity concerns the fact that there is frequently not a linear relationship between how much an organizational unit does of a given activity and the

[1] In the 'narrow' definition of 'productivity equals efficiency'.

contribution of that amount to the productivity of the unit (e.g. a point of diminishing returns). These *non-linearities* should be represented in the measurement system (Pritchard & Roth, 1991).

It was noted above that both measures of *efficiency and effectiveness* should be used in the motivational approach toward productivity. The measurement system should therefore be able to accommodate both types of measures.

A productivity measurement system should take into account *interdependencies* between individuals and units. The majority of work in an organization is done interdependently, which makes it difficult if not impossible to identify the contributions of individuals to the team effort. It is the exception rather than the rule that the productivity of a unit is a simple summation of the productivity of its individual members. This issue will be discussed in more detail in Section 2.2.5.

Another criterion is *flexibility*. Organizational goals and policies may change over time. The productivity measurement system must be able to incorporate such changes without much difficulty.

A final criterion is an essential one: the system should be *accepted* by the organization (incumbents, supervisors, and management). The best, most valid, system imaginable would be ineffective if the organization did not accept it and find it useful.

When productivity measurement is an organization-wide effort, two additional features are desirable. Firstly, the system should be able to aggregate the measurement systems of different units into a single broader system. Secondly, it would be very useful if the productivity of units with different functions could be directly compared.

A number of productivity measurement systems confirm to at least the first four requirements (overall index, subindices, validity, efficiency and effectiveness measures). Among these are 'Methodology for generating efficiency and effectiveness measures (MGEEM)' (Riggs & Felix, 1983; Tuttle & Weaver, 1986a, 1986b) and the 'Multicriteria performance/productivity measurement technique (MCP/PMT)' (Sink, 1983, 1985). In principle, the 'Productivity Measurement and Enhancement System (ProMES)', developed by Pritchard and his associates (Pritchard, 1990; Pritchard et al., 1986, 1987, 1988, 1989), and described in Section 1.4, meets all criteria mentioned above.

2.2. PERFORMANCE MANAGEMENT

The concepts of goal setting and feedback have become central to managing human performance. Their effectiveness in enhancing performance has been amply demonstrated in both laboratory and field studies (e.g. Kopelman, 1986; Locke & Latham, 1990). Goal setting and feedback are also central elements of most performance management programs aimed at improving personnel productivity. These programs have been very successful (Guzzo et al., 1985).

In this section, an concise overview will be given of the theories of goal setting and feedback. Some main findings regarding effects on performance will be discussed. Attention will be given to the way goals and feedback interact in producing these effects. Also, demands posed by task complexity and task interdependence are discussed.

2.2.1. Goal setting theory

One of the most prominent theories of work motivation is the goal setting theory, which resulted from almost 30 years of empirical research. The core premise of goal setting theory is that goals are immediate (though not sole) regulators of human action. Research on goal setting (reviewed in Locke & Latham (1990), which is the main source for this section) resulted in two main findings regarding effects on performance. These will be discussed briefly, as will be the mechanisms through which goals affect performance, factors influencing commitment to goals, the role of self-efficacy, and the moderating effect of task complexity on the goal-performance relationship.

Effects on performance: main findings

First, there is a positive linear relationship between goal difficulty and performance. This means that difficult goals lead to higher task performance than easy goals (with moderate goals in between). Only when subjects reach the limits of their ability at high goal difficulty levels, does this function level off. The explanation for this goal difficulty effect is that difficult goals lead to greater effort and persistence than easy goals, assuming that the goals are accepted.

Second, goals that are specific and difficult lead to a higher level of performance than vague, non-quantitative goals such as "do your best", or no assigned goals. An explanation for this finding is that vague "do-your-best"-goals allow individuals to give themselves the benefit of the doubt in evaluating their performance, since a wide range of performance levels may be interpreted as fulfilling the criterion of doing one's best. In contrast, specific, hard goals define success as attaining a certain high score. Goal specificity, viewed separately from goal difficulty, has the effect of reducing performance variance, because it reduces interpretive leeway as to the exact meaning of the goal (Locke, Chah, Harrison & Lustgarten, 1989).

Goal mechanisms

Goals have their effect on performance through four mechanisms (Locke and Latham, 1990). First, goals energize performance by motivating individuals to exert *effort* in line with the difficulty of the goal or task. Second, goals motivate individuals to *persist* in their activities through time. Difficult goals ensure that an individual will keep working either for a longer period of time or, where time limits are imposed, faster or harder than would be the case with vague or easy goals. Third, goals, especially if they are specific, *direct attention* to activities that are relevant for goal attainment at the expense of activities which are not goal-relevant. These three mechanisms–effort, persistence, and direction of attention–are relatively direct and automatic consequences of goal-directed activity. They start to operate automatically once the individual commits to a goal and decides to act to achieve it. A fourth mechanism is strategy development; this may come into operation if more intensive execution of 'automated' task strategies is not sufficient for goal attainment. An individual may then try to develop better task-specific strategies for performing the task. This involves conscious problem-solving and creative innovation. Task strategies may also be developed as a means of saving effort (e.g. "work smarter rather than harder").

Goal commitment

Goal commitment (or: goal acceptance) refers to the extent to which an individual is attached to a given goal, considers it significant or important, is determined to reach it, and keeps it in the face of setbacks and obstacles (Latham & Locke, 1991). So, a goal to which a person is not committed is not really a goal and therefore cannot have much effect on subsequent actions and performance. Two categories of factors affect goal commitment (Locke & Latham, 1990). There are factors which affect the perceived desirability of trying for a given goal. These include leader authority (e.g. supportiveness, trust, physical presence, pressure, provision of rationale for the goal), peer group pressures, public commitment to goals, and incentives and rewards. The second category refers to factors affecting the perceived ability of attaining a given goal. This category includes ability, experience, training, information about appropriate task strategies, past success, and internal attributions.

The factors that affect goal choice are similar to those affecting goal commitment mentioned above. The most direct method of influencing goal choice is for an authority figure to assign the goals. Locke and Latham (1990) report a correlation between assigned goals and subsequently self-set goals of .50.

Goals, self-efficacy, and performance

Self-efficacy is defined as a judgement of how well one can execute courses of action required to deal with prospective situations (Bandura, 1982). It could be labelled "task-specific self-confidence". Self-efficacy has powerful direct effects on performance. In addition, self-efficacy can affect performance indirectly by influencing goal choice and commitment to goals. So, performance is affected not only by what one is trying to do but also by how confident one is of being able to do it (Locke & Latham, 1990). Self-efficacy is influenced by enactive mastery (actual performance or beliefs about performance), role model performance, persuasion (convincing subjects that they will be able to perform well or that performance on the task is controllable), and providing task strategy information (Locke & Latham, 1990). In this respect, supportive leadership may enhance self-efficacy and lead to the setting and acceptance of more difficult goals than non-supportive leadership behavior (Latham & Saari, 1979a; Locke & Latham, 1990).

Goals, task complexity, and performance

Task complexity[2] has been shown to have a significant moderating effect on the performance gains that result from specific, difficult goals (Wood, Mento & Locke, 1987). The effects of setting specific difficult goals are typically larger on easy tasks (reaction time, brainstorming, simple arithmetic) than on complex tasks (business game simulations, technician work, scientific and engineering work). This difference may be caused by the different mechanisms that come into operation on simple and complex tasks. On simple tasks, goals affect performance directly through one or more of the three automatized universal task strategies (direction of attention, effort and persistence) and one or more automatized task-specific strategies (plans). As task complexity increases, all of these mechanisms become less adequate to ensure goal achievement, while problem-solving and the development of task-specific strategies become more important (Wood & Locke, 1990).

2.2.2. Feedback theory

Feedback about the effectiveness of behavior has long been recognized as essential for motivation and learning in performance-oriented organizations (Ilgen, Fisher & Taylor, 1979). In this section, some cognitive and motivational effects of feedback are discussed.

Effects on performance

There is evidence that individuals prefer specific, timely and positive feedback (e.g. Ilgen et al., 1979) and that such feedback enhances performance (e.g. Kopelman, 1982, 1986).

Feedback (or: knowledge of results) affects task performance in two ways. First, feedback improves task performance for cognitive reasons (Kopelman, 1982). Feedback corrects misperceptions, it can provide information about the correctness, accuracy, and adequacy of work behaviors and it can instruct people as to the specific behaviors that should be performed and/or the specific outcomes that should be achieved. So, objective feedback reduces role ambiguity and causes appropriate behaviors and/or outputs. Evidence has repeatedly shown that the effects of feedback are limited to the specific behaviors or outputs for which feedback is provided. Therefore, the more specific the feedback is, the greater the effects are[3].

---

[2] Three task characteristics contribute to task complexity (Wood, 1986): component complexity (the number of acts and information cues that are inputs to a task product), coordinate complexity (the relationship between task inputs and task products, e.g. sequencing of acts, time allowed), and dynamic complexity (changes in the acts and information cues).

[3] Specific feedback facilitates the setting of specific goals, which have larger effects on performance than general 'do-your-best-goals'.

The second major reason why objective feedback improves task performance is because it can enhance motivation[4]. For example, feedback may be necessary for instilling a sense of competence, accomplishment and control in individuals (e.g. Hackman & Oldham's Job Characteristics Model, 1976, 1980). Feedback can also increase motivation to the extent that it creates social consequences. For instance, public feedback may increase competition among people, as they vie for respect and recognition and try to avoid embarrassment that may accompany poor performance (e.g. Runnion, Watson & McWhorther, 1978). Furthermore, feedback may highlight potential external consequences. For example, whereas the absence of performance management implies that management finds any level of performance to be acceptable (Law, 1975, cited in Kopelman, 1982), by instituting performance management and feedback, management conveys an implicit message that it is interested in the individual's or group's performance and that certain performance levels may have more or less favorable consequences.

Basically, two types of feedback can be distinguished: outcome feedback (information concerning performance outcomes) and process feedback (information concerning the manner in which an individual implements a work strategy). Both types of feedback interact with goal setting to enhance performance. The effects of both types of feedback may be additive according to Earley, Nothcraft, Lee & Lituchy (1990), who found that the combination of specific, challenging goals, and both specific process and outcome feedback produced a higher level of performance than other combinations.

Feedback, task complexity, and performance

In novel, complex, or unstructured tasks in which the relation of behaviors to outcomes may be uncertain, feedback focusing on the behavioral processes that generate outcomes may be more beneficial than outcome feedback, since it may increment knowledge about the task or helps individuals understand task performance (Earley et al., 1990; Ilgen et al., 1979; Jacoby, Mazursky, Troutman & Kuss, 1984; Steele Johnson, Perlow & Pieper, 1993).

---

[4] While everyone agrees that feedback may increase performance through its cognitive/informational properties, there is considerable disagreement in the literature about whether goals are a necessary complement to feedback to improve performance through increased motivation (e.g. Locke & Latham, 1990; Nadler, 1979). In general, feedback in absence of formal goal setting will often cause individuals to engage in spontaneous informal goal setting, which leads to the question whether it is not in fact the setting of goals which causes the motivational effects. If one differentiates between explicitly formulated and publicly announced goals and intentions that are less clearly stated (Algera, 1990), a large part of the motivational effects of feedback seem to occur through attainment of latter (implicit) goals. However, it is also clear that goal setting without feedback is hardly effective (Locke & Latham, 1990). The interaction between goal setting and feedback is discussed in Section 2.2.3.

2.2.3. The interaction of goal setting and feedback

Although both goal setting and feedback have been demonstrated to have positive effects on performance and proponents of one concept tend to downplay the importance of the other concept, it should be noted that neither is very effective in the absence of the other. Locke & Latham (1990) define the relationships between goals and feedback as follows: "With respect to feedback, goals are a *mediator*; they are one of the key mechanisms by which feedback gets translated into action. With respect to goals, feedback is a *moderator*; goals regulate performance more effectively when feedback is present than when its absent." Research indicates that the combination of goal setting and feedback is consistently more effective than either feedback or goal setting alone (Balcazar, Hopkins & Suarez, 1986; Locke & Latham, 1990).

When introducing formal feedback without formal goal setting, spontaneous goal setting can be expected to occur (e.g. Bandura & Cervone, 1983). Conversely, if goal setting is introduced without formal feedback, feedback seeking behavior is likely to occur (e.g. Ashford, 1986). So, the absence of formal feedback or goal setting procedures does not mean that no goals are set or no feedback is sought.

The interaction of goals and feedback in regulating performance involves numerous complex cognitive processes within a control systems model (Campion & Lord, 1982; Locke & Latham, 1990). The response to feedback in relation to initial goals is affected by the degree and direction of the discrepancy from the initial goal (Matsui, Okada & Kakuyama, 1982), the amount of dissatisfaction with the performance (Matsui, Okada & Inoshita, 1983), and self-efficacy (Podsakoff & Farh, 1989).

When a goal is set and subsequent feedback is provided, performance depends on a number of appraisals and decisions that follow (Locke & Latham, 1990). If there is a low discrepancy between the goal and the feedback, an individual will usually be satisfied and try to maintain the same level of performance. However, attaining previously mastered performance levels may become less satisfying, causing the individual to raise his or her goal. If there is a high goal-performance discrepancy, but the individual is not dissatisfied (e.g. because he/she thinks the goal is not very important), performance will tend to stay the same. If the individual, however, is dissatisfied with the performance level, what will happen will depend largely on the individual's self-efficacy. If the person has high self efficacy, he/she will tend to set high goals, which will lead to increases in motivation and performance (assuming adequate ability). If the individual has low self-efficacy, he/she will tend to set lower goals, which result in limited increases or even decreases in performance.

2.2.4. Reinforcement

Reinforcement refers to the application of Skinner's techniques of operant conditioning to the control of behavior. The premise of reinforcement is that behavior is largely a function of its consequences. If the consequences of a behavior are positive, the probability that the behavior will be repeated increases. On the other hand, if the consequences are negative, the probability that the behavior will be repeated decreases (e.g. Latham & Wexley, 1981; Luthans & Kreitner, 1975). This finding is, among other things, important when setting goals: goal acceptance depends to a large extent on whether the individual perceives goal attainment as instrumental in providing positive consequences.

There are three types of behavioral consequences: reinforcers (positive or negative), punishers, and neutral events. A positive reinforcer is a consequence of behavior (e.g. praise, a reward) that increases the probability that the behavior (e.g. producing high-quality output) will be repeated. A negative reinforcer is a consequence (e.g. nagging), the termination of which immediately after a behavior (washing the dishes) started to occur increases the probability that the behavior (washing the dishes) will be repeated. The probability of repeating a behavior is decreased by a punisher (e.g. a speeding ticket–hopefully–reduces a driver's tendency to exceed the speed limit). An important problem in using punishment in an organizational setting is that people tend to avoid punishment and therefore the punisher, which decreases the effectiveness of leadership. Interestingly, the absence of reinforcement–ignoring the behavior–also *decreases* the probability of the behavior being repeated ('extinction').

Four basic principles apply to the use of reinforcers (Latham & Wexley, 1981, in a performance-appraisal context). First, the reinforcer must be made contingent upon the desired behavior. Secondly, the individual must clearly perceive the relationship between the desired behavior and the reinforcer. Thirdly, the reinforcer must be administered soon after the desired behavior has been emitted. And lastly, the reinforcer must represent a valued outcome for the employee.

2.2.5. Task interdependence

Although the majority of goal setting and feedback studies have used the individual as the unit of analysis, the studies that focused on groups show nearly the same success rate (Locke & Latham, 1990). Research on the effects of goals and feedback in interdependent tasks (tasks where individuals have to work together to some degree to accomplish the task) suggest that a combination of individual and group feedback may be most effective, whereas only providing individual feedback may be dysfunctional. Matsui, Kakuyama & Onglatco (1987), studying a group task in which members were highly interdependent, found that effectiveness of task feedback is maximized when it provides information on both group and individual performance. This finding is explained using the control systems model of task feedback (Campion & Lord, 1982) in which a negative discrepancy between the goal/target and the feedback provides the impetus for discrepancy reduction and thus performance improvement. If feedback does not include individual feedback, those individuals below target will not improve performance if the group performance is on target. On the other hand, if feedback does not include group feedback, those individuals whose group is below target will not improve their performance if they themselves are on target. Only if feedback includes both group and individual task feedback, does information of lower group (individual) progress trigger increased effort to improve performance, despite feedback suggesting higher individual (group) progress. In this way, utilization of individuals' resources is maximized. Furthermore, motivation loss due to social loafing is minimized. Social loafing (e.g. Latané, Williams & Harkins, 1979) refers to the tendency for individuals to 'free-ride' on the efforts of other group members. If latter group members find their partners free-riding on them, they may lose motivation (the so-called sucker-effect). Motivation loss (and an accompanying performance decrease) due to social loafing can be minimized by decreasing dispensability of efforts or by increasing task identifiability. The former is done by setting both group and individual goals. The latter is done by providing feedback in relation to both goals. Mitchell & Silver (1990) found that setting individual goals in an interdependent task resulted in lower performance than setting group goals, group goals plus individual goals, or no goals at all because of dysfunctional effects resulting from competitive feelings, strategies and behavior resulting from individual goals.

Regarding distribution of rewards among group members under high interdependence, the general conclusion is that reward systems that distribute payoffs equally (low differentiation) result in more cooperation and higher performance than reward systems in which some group members get a much higher reward than others (high differentiation) (e.g. Rosenbaum, 1980).

2.3. PARTICIPATION

A key characteristic of the ProMES method is the participative development of the productivity measurement system. One of the main research questions of this study is whether or not this participative development is an essential precondition for commitment to the system and improved productivity. In this section, we will therefore give some theoretical and empirical background on the concept of participation, the effect of participation on productivity and satisfaction, the mechanisms through which these effects occur, and some moderating factors.

2.3.1. Defining participation

A general definition of participation (Bass, 1981) is: "a process of sharing in some activity between and/or among superiors and subordinates". Somewhat more specific definitions are: "joint decision making by two or more parties" (Vroom, 1960) and "influence-sharing between hierarchical supervisors and their subordinates" (Mitchell, 1973). Participation implies relinquishing an authoritarian, directive form of leadership and accepting a form of decision-making that gives subordinates opportunities to influence the decision (Van der Vlist, 1990).

Usually, participation is presented as a continuum, with at one extreme decision-making exclusively by the leader and at the other extreme decision-making by the group, with the leader being one of the members who at most specifies criteria or constraints the decisions should meet (Van der Vlist, 1990). Some authors include 'delegation' at this extreme of the continuum. For example, Heller (1971) considers five types of decision processes along this continuum: (1) own decision without explanation, (2) own decision with explanation, (3) prior consultation with subordinate(s),

(4) joint decision-making with subordinate(s), (5) delegation of decision to subordinate(s).

2.3.2. Effects and mechanisms: introduction

The effects of participation can be divided into two categories (Locke & Schweiger, 1979). The first category concerns aspects of task performance and productivity such as better decision-making, higher output, lower costs, or better quality. The second category concerns job satisfaction and morale and their concomitants, such as reduced turnover, absenteeism, and conflict.

Three main models propose mechanisms through which participation influences performance and satisfaction: the cognitive, the motivational, and the contingency model. These models are not mutually exclusive and variables from all three models play important roles in the participative process (Miller & Monge, 1986).

In *cognitive models* participation is recommended because it enhances the flow and use of important information in organizations. Because workers have more complete knowledge of their work than management, high quality information is brought to decisions by having these workers participate. In addition, employees will know more about the rationale for and the implementation of the decisions if they participated in the decision-making process (e.g. Pritchard, 1990)

Motivational models suggest that participation has its effects through such factors as increased trust, greater control of the work, more ego involvement in the job, increased identification with the organization, more group support (in case of group participation), and the setting of higher goals and increased goal acceptance (Locke, Schweiger & Latham, 1986).

Proponents of the *contingency models* suggest that participation will affect satisfaction and productivity differently for different people and situations. Contingency models predict that personality, situational influences and values mediate the effects of participation on productivity and satisfaction. A well-known elaboration of the contingency model is the work of Vroom (e.g. Vroom & Yetton, 1973) who proposes that the appropriateness of forms of participation is dependent on the required quality and acceptance of a decision, on the complexity of the situation at hand, and on time constraints.

In the remainder of this section, a distinction will be made between *participation involving goal setting (PGS)*, and *other forms of participation in decision-making (PDM)* which are not part of a formal goal setting process.

2.3.3. Participation in decision-making (PDM)

Effects on task performance

Recent review studies on the overall effects of PDM show either small positive effects on performance or no effects at all. Results from a meta-analysis by Miller and Monge

(1986) showed a small positive correlation between PDM and 'productivity' (r=.16). Locke and Schweiger (1979) reviewed laboratory and (correlational and controlled experimental) field studies concerning the effects of participation. For neither setting or type of study did they find evidence to suggest that PDM is consistently superior or inferior to non-PDM. PDM was superior to non-PDM in 23 percent of the studies, no difference was found in 51 percent of the studies and PDM was inferior to non-PDM in 26 percent of the studies. These results were confirmed by Schweiger and Leana (1986) who found 24, 42, and 24 percent respectively. A re-analysis of Locke and Schweiger's results by Van der Vlist (1990) resulted in a more positive assessment of PDM (28 percent superior, 67 percent no difference, 5 percent inferior).

Andriessen and Drenth (1984) assume that the reason for the lack of clear effects of participation on motivation, performance, or productivity is that the relation between the participation concept, which varies considerably according to the situation, and the performance concept, which is determined by numerous factors, is very complex and largely dependent on the situation.

Some cognitive and motivational mechanisms that may cause increases in performance are the following (listed in Locke & Schweiger, 1979). From a cognitive viewpoint, PDM may result in increased information, knowledge, and creativity which helps in solving organizational problems through better upward communication and better utilization of knowledge. PDM may cause better understanding on the part of the employees who are to execute the decisions resulting from PDM. With regard to motivation, PDM may result in less resistance to change because of increased trust on the part of the employees and/or a greater feeling of control and reduced anxiety. Acceptance of and commitment to decisions and changes (including goals, see PGS), may increase through greater degree of ego involvement or identification with the organization and through the effects of group pressures.

Effects on satisfaction

There is somewhat stronger evidence for the relationship between participation and satisfaction than for that between participation and productivity. Based on their meta-analysis, Miller and Monge (1986) found a positive correlation of .34 between participation and satisfaction. Locke and Schweiger (1979) found PDM to be superior to non-PDM in a large percentage of both laboratory and field studies. This was the case in 65% of the studies they reviewed, whereas there is no difference in 27% of the studies, and only 8% of the studies reported PDM to be inferior to non-PDM in causing satisfaction. Schweiger and Leana (1986) report very similar results (68, 23, and 9 percent respectively).

According to Locke and Schweiger (1979), the simplest explanation why participation leads to increased job satisfaction and morale is that allowing participation will increase the likelihood that employee will get what he wants (e.g. Mitchell, 1973), that he will attain his values. Whether value attainment will happen through participation depends on what the employee wants (e.g. simply express his views, actual influence on decision-making, or full equality with his supervisor) and whether PDM facilitates this.

Participative development of ProMES: a form of PDM

In discussing the reasons "why ProMES works", Pritchard (1990) stresses the importance of having the personnel who are going to be using the system heavily *involved* in its development. Unit personnel often consider programs imposed from above ineffective, because they feel they were designed without taking into account their unique needs and environment. They feel that the people designing these systems do not have a full understanding of the work of the unit. It is more effective to have heavy involvement from the unit personnel so that the final system will fit their work, and they will not feel that it is another project imposed on them from above. This also helps with the *acceptance* of the program and reduces the likelihood that personnel will ignore or sabotage it. Unit personnel identify the products and indicators, they develop the contingencies, and they defend their work to management. Hence they have a sense of ownership.

According to Pritchard (1990), the system also improves *role clarification*. The process of developing, refining and getting approval for the products, indicators, and contingencies helps personnel understand their roles more clearly. Unit personnel discuss what their objectives should be, disagreements surface and are resolved. Expected levels of output are discussed and consensus is achieved. When the development process is finished, the units have a much clearer picture of what their objectives are, what they should focus on to achieve their objectives and what is expected of them in each area. This role clarification process should have motivating properties in itself.

It is clear from the above that both the motivational and the cognitive model are represented in Pritchard's preference for a high degree of participation in developing a ProMES system.

Possible moderating factors

Many moderating factors have been suggested for the relationship between PDM and performance and satisfaction (e.g. Andriessen & Drenth, 1984; Campbell & Gingrich, 1986; Locke & Schweiger, 1979; Schweiger & Leana, 1986; Wagner & Gooding, 1987). Some individual factors (knowledge, motivation) and some situational factors (task complexity, group characteristics, leader attributes, time pressure, and organizational change) will be discussed briefly.

Knowledge. PDM should be most helpful in generating high quality decisions when the participants have relevant knowledge to contribute. In case where one member (e.g. the leader) has significantly more knowledge than the others, PDM would be a waste of time and effort, or even harmful to decision quality (in case those with less knowledge outvote the most knowledgeable member).

Motivation. PDM may not satisfy employees who do not want or expect PDM, who lack independence and want to be told what to do. Furthermore, PDM may not be effective with employees who are not used to it, though employees might get used to PDM if it is used frequently. Finally, although it has been asserted that PDM may be less effective with generally less motivated employees, research also implies that PDM may enhance motivation, because it gives those employees a feeling of control and higher efficacy.

Task complexity. Highly complex unstructured tasks may require PDM, because of the increased knowledge and flexibility requirements, whereas routine tasks do not. However, PDM may enrich simplified work, so that employees doing routine jobs may become more committed as a result of PDM, while those at higher level jobs may not need no such incentive. For example, Wagner & Gooding (1987) found that PDM enhanced satisfaction and acceptance on simple tasks significantly more than on complex tasks.

Group characteristics. PDM may increase group conflict. Although extreme conflicts probably have negative consequences, constructive use of differences of opinion may lead to higher decision quality than does the absence of conflicting viewpoints. On the other hand, PDM may lead to group conformity and/or groupthink (e.g. Janis, 1972).

Group size seems to be a moderator too: an increase in group size not only affects the time needed to reach a decision, but also increases the problem of regulation and coordination. These problems may be solved by using methods which reduce the amount of face-to-face interaction, such as the Nominal Group Technique and the Delphi Technique (Delbecq, Van de Ven & Gustafson, 1975).

Leader attributes. Supervisors themselves are often threatened by the introduction of PDM and oppose and try to sabotage it (e.g. Scheflen, Lawler & Hackman, 1971). Any successful PDM program should therefore have the support of all levels of management. When using PDM, the skills of the group leader in using the PDM techniques effect their success. Also, the personality and style of the leader in relation to that of the group may also be important. When leader and group have similar values, the members may be more satisfied.

Time pressure. PDM is more time consuming than non-PDM. So, if there is pressure for an immediate decision, PDM may not be suitable.

Organizational change. When there is rapid or constant organizational change involving complex knowledge requirements which cannot be mastered by a single person, PDM should be superior to unilateral decision making. PDM should also facilitate change when the changes are threatening to employees if it gives them a heightened sense of control (Locke & Schweiger, 1979)

2.3.4. Participation in goal setting (PGS)

Effects on performance

Motivational effects. Locke & Latham (1990) reviewed 11 studies (four laboratory studies, seven field studies) into the effects of participation in goal setting (PGS) on performance. They found that it is the specificity and difficulty of the goal that is set that affects performance rather than whether the goal is assigned or set participatively. The effects of assigned goals are as powerful as participatively set goals in generating high goal commitment and subsequent performance. More specifically, Latham & Steele (1983) conclude:

- specific goals lead to higher performance than do general 'do best' goals;
- there is a linear relationship between goal difficulty and performance;
- participation in task strategy, from a *motivational* point of view, affects performance only to the extent that it includes the setting of a specific goal and/or a specific goal that is more difficult than the one that is assigned unilaterally by the supervisor.

Latham, Erez & Locke (1988) found that assigning goals in a "tell and sell" manner (i.e. friendly, supportively, stressing that the goal is attainable) is indeed as effective as setting the goal participatively. Assigning the goal in a "tell" manner (very brief, unsupportive) is less effective than either participatively setting the goal or assigning the goal in a "tell and sell" manner.

Erez (Latham, Erez & Locke, 1988) mentions the fact that in real life situations there are many cases in which employees are required to change previously set goals, or to follow goals that are not in line with their personal aims. In such cases, employees are more likely to reject the goals assigned to them. She, therefore, concludes that PDM is most effective when the situational characteristics are the least favorable for goal commitment.

In the same article, Locke notes that participation in the absence of self-efficacy may not lead to high performance and could even lead to increased stress, in that people will be faced with the need to cope with situations they cannot handle. Thus procedures that increase subject choice should be most successful when combined with additional procedures that promote self-efficacy with respect to the task in question.

Cognitive effects. Campbell & Gingrich (1986) studied the interactive effects of task complexity and participation on task performance, focusing on the cognitive implications of participation. They conducted a field experiment with 40 computer programmers writing either a simple or complex program. Half the participants were actively involved in discussing the project and in jointly determining the completion target. The other half were simply assigned completion times of equivalent difficulty. The findings demonstrate that the performance on a *complex* task can be significantly improved through participation. The authors presume that this improved performance comes about due to a cognitive mechanism: the generation of more information and better–clearer, more focused–information in the participation process. During the process, the supervisor can become aware of gaps or misconceptions in the subordinate's understanding of the task and can then provide clarifying information. Similarly, the person carrying out the task can explore the possible causes of action with the supervisor, identifying and eliminating approaches considered unrealistic or unacceptable.

Effects on goal acceptance/commitment

Similar to the performance results, the largest percentage of the laboratory and field studies found no differences between participatively set and assigned goals on goal acceptance/commitment. Schweiger & Leana (1986) reviewed 11 studies (five laboratory studies, six field studies). In only one laboratory study did PGS lead to higher goal acceptance than assigned goal setting. One field study found PGS to be inferior to assigned goal setting. One field study (Latham & Yukl, 1975) reported subgroup differences: in one group (uneducated logging crews) PGS was superior, in the other group (educated logging crews) there were no differences. In eight studies, no differences in goal acceptance were found.

Effects on goal difficulty

According to Locke & Latham (1990), PGS will have a positive effect on task performance if it causes higher goals to be set than would be the case if the goals had been assigned. Very few studies have investigated the effects of participatively setting a goal versus assigning the goal on goal difficulty (many PGS studies have controlled goal difficulty rather than investigated its effects directly). Of the five studies mentioned in Schweiger & Leana (1986), one field study showed a positive effect of PGS on objective goal difficulty, three studies reported no differences. Latham & Yukl (1975) found subgroup differences: for an uneducated group, participatively set goals led to greater goal difficulty than assigned goals, whereas there were no significant differences for an educated group.

2.3.5. Summary

To summarize, whereas participation in decision-making (PDM) usually has a moderately positive effect on satisfaction, effects on performance are usually smaller or nonexistent. Several individual and situational factors may moderate the relationship between PDM and productivity and satisfaction. Participation in goal setting (PGS) will only have positive motivational effects on performance if participation results in the setting of more difficult goals than assigning the goal. An important moderator of the PGS-performance relationship appears to be task complexity: on complex tasks, PGS may cause improved performance through cognitive effects.

2.4. A NORMATIVE MODEL

2.4.1. Introduction: the accepted control loop

A central concept in this thesis is the 'accepted control loop' (Kleingeld & Van Tuijl, 1992, in press). This concept will be explained using Figure 2.1, which is an abstract representation of controlled processes in organizations. It shows the transformation of inputs into outputs, which is controlled by a regulator. To achieve an optimally controlled process, the regulator should have goals (standards) that tell him how the transformation process should perform. These goals pertain to the utilization of inputs in the transformation process (efficiency), to the transformation process itself, and to demands on the outputs of the transformation process (effectiveness). Furthermore, the regulator should have the ability to intervene in the process. Finally, the regulator needs feedback from the input-transformation-output process, e.g. information about the use of inputs in the transformation process, the operations performed in the transformation process, and the quality of outputs. Only discrepancies between accepted feedback and accepted goals will cause a regulator to take action aimed at reducing the discrepancies and thus attaining the goals.

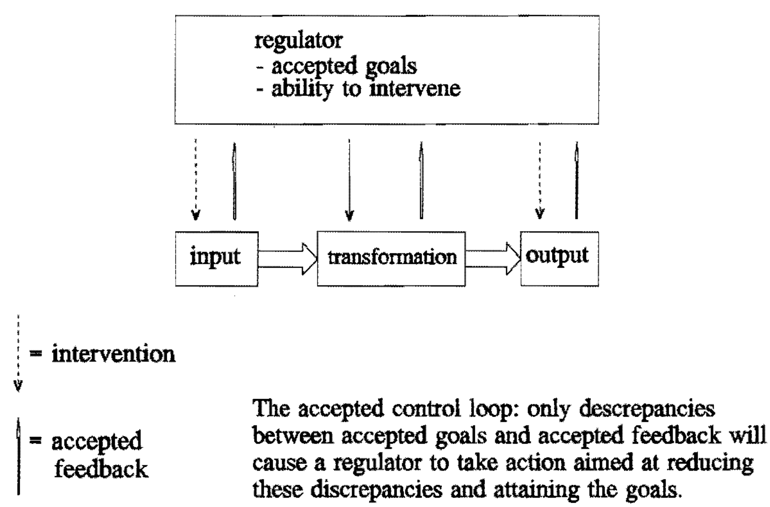


Figure 2.1. The accepted control loop.

ProMES can be considered a method of designing an accepted control loop with which an organizational unit regulates its performance. A number of design criteria should be fulfilled for the ProMES control loop to be effective. The literature on goal setting and feedback supplies general criteria for an effective control loop. The literature on ProMES supplies guidelines on the design process and the characteristics of the

ProMES system and its implementation. Finally, the organizational context poses constraints that have to be taken into account in designing and implementing the control loop. In this study, these criteria and constraints were used to guide the design and implementation of the ProMES system and as a model for analyzing problems that occurred. Figure 2.2 provides an overview of these criteria, discussed in the remainder of this section.

2.4.2. General design criteria

Setting performance goals and providing feedback is considered one of the most powerful psychologically-based techniques for improving the task performance of individuals and groups. Based on the literature (summarized in Section 2.2), the following criteria are considered essential for effective application of goal setting and feedback[5], in that their attainment facilitates the comparison of accepted feedback and accepted goals, which is an central element of the 'accepted control loop'.

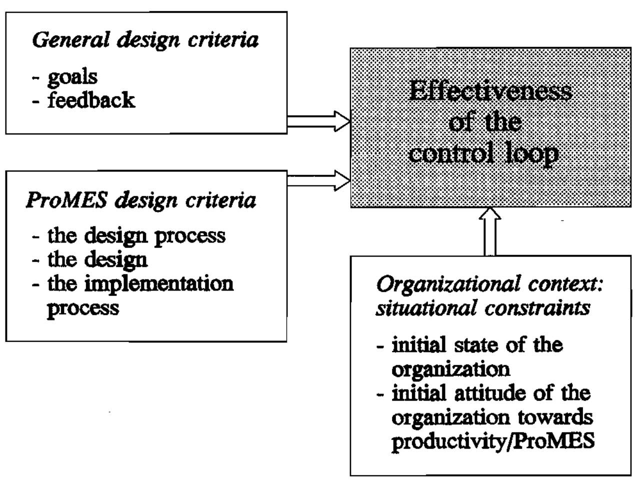


Figure 2.2. Design criteria and situational constraints.

Goal criteria

Challenging goals. Effective goals are challenging, i.e. difficult but attainable. Difficult goals lead to greater effort and persistence than easy goals, assuming that the goals are accepted. The goals set should not, however, be so difficult that they are (perceived as)

[5] No design criteria will be generated for the reinforcement concept. However, the importance of consistent reinforcement is reflected in one of the context factors, viz. performance appraisal and rewards.

unattainable, since that may cause rejection of the goals. In addition, supportive leadership is recommended, because it may lead to the setting and acceptance of more difficult goals than non-supporting leader behavior.

Specific goals. Effective goals are specific, i.e. goals that are specific and challenging lead to a higher level of performance than vague but challenging goals such as "do your best," vague but unchallenging goals such as "work at a moderate pace," or the setting of no goals. A specific challenging goal clarifies what constitutes effective performance. Consequently, a person is no longer able to interpret a wide range of performance levels–including those that are lower than the person's actual best–as indicative of excellent performance.

Goal controllability. Whether a goal is attained should be determined by the direction, effort and persistence of the individual's (or group's) behavior. These are internal, controllable factors (Klein, 1989). The influence of external, uncontrollable, factors should be limited. The point is that individuals will develop causal explanations for being unable to meet their goals (Carver & Scheier, 1981); if the failure to attain the goal is attributed to external factors, especially when they are stable, and the individual believes no behavioral response can remedy the situation, the likelihood of remaining committed to the goal will decrease (Taylor, Fisher & Ilgen, 1984).

This issue of goal controllability ties in directly with the controllability criterion for performance indicators (a ProMES design criterion in the following section). The indicators used for goal setting should exclude external, uncontrollable factors as much as possible.

Completeness of goals. Goal setting is a very powerful motivational tool. Specific, difficult goals have a large effect on the direction, effort and persistence of behavior and thereby on performance *on the task dimension for which the goal is set* (Locke & Latham, 1990). This implies that specific, difficult goals should be set for all important aspects of task performance. For example, if both quantity and quality of production and safe work practices of a unit are important for the company, specific, difficult goals should be set on all three dimensions of task performance (even though it might be more difficult to develop specific and difficult goals for quality or safety).

Feedback criteria

(Perceived) validity/accuracy of feedback. The feedback should be an accurate representation of task performance. This means that feedback is generated through valid measures of task performance on relevant performance dimensions. Employees' behavior should be sampled adequately before giving feedback (Taylor et al., 1984). Findings from several studies have indicated that individuals assess the accuracy of feedback they receive and that the perceived accuracy of this information moderates their reactions to it (Taylor et al., 1984). Since feedback perceived as inaccurate receives little weight in individuals' overall assessments of their current behavior, it is

important that the recipients view the feedback as accurate. The perceived accuracy of the feedback depends, amongst others, on perceptions of the expertise of the feedback source, on the recipient's trust in the source's motives, and on the consistency of the feedback (Ilgen et al., 1979). The perceived accuracy of the feedback may be enhanced by specific support for the feedback (e.g. critical incidents). Having the future recipients participate in the development of the feedback measures (an essential part of the ProMES method) may lead to positive perceptions of feedback accuracy (Pritchard, 1990).

Completeness of feedback. Feedback should be provided on all important performance dimensions or goals. Goals regulate performance far more reliably when feedback is present than when it is absent. When feedback is withheld from individuals with goals, goal setting seems to have little effect on performance (Locke & Latham, 1990). Effects of feedback are limited to the specific focal behaviors or outputs for which feedback is given (Kopelman, 1986). So, providing feedback on a subset of the important performance dimensions or goals (for instance because they are the ones on which feedback information is readily available) can have detrimental effects, since the dimensions not measured will be neglected in favor of the dimensions measured ("What you measure is what you get", Pritchard, 1990).

Understandability and surveyability of feedback. For effective processing of feedback to occur, it should be easily understood by those at whom it is aimed. It should be adapted to the level of comprehension of the employees. Graphical representation of feedback may have a much more substantial impact on employee behavior than merely feeding back raw data, because it helps them get a quicker understanding of both the details of the situation and the big picture (Connellan, 1978).

Timeliness of feedback. In general, the shorter the interval between an individual's behavior and the receipt of feedback about the (results of the) behavior, the better (Connellan, 1978). Deviations from a performance standard become quickly apparent and the appropriate corrective actions can be taken. This holds especially true if the time period between the behavior and the feedback is filled with activities which interfere with the individual's ability to accurately recall the behavior and associate the feedback with it (Ilgen et al., 1979). The feedback frequency should be attuned to the job cycle (Pritchard, 1990).

Specific feedback. Feedback should provide specific information on progress in relation to a goal and avoid generalizations that do not provide precise information ("You did an O.K. job last month") (Connellan, 1978). Because goals focus attention on information that is considered to be significant and direct subsequent action with respect to it, they influence the perceived importance of feedback. Information that is not linked to goals will be relatively low, and will therefore not lead to action (Latham, Mitchell & Dossett, 1978; Locke & Latham, 1990).

Positive feedback. Positive feedback (indicating that the goal has been attained) is pleasant and may enhance one's self-image and, therefore, tends to be accepted more readily than negative feedback (indicating that the goal has not been reached). Negative feedback may be rejected as inaccurate by the recipient because of an unwillingness to accept such knowledge about himself or herself (Ilgen et al., 1979). To diminish negative reactions in the case of negative feedback, it is advised to present the feedback in a descriptive form (focusing on the positive elements as well as on the negative ones), to include specific behaviors, to set difficult but attainable goals, and to 'stress the positive' as much as possible (e.g. Latham & Wexley, 1981).

2.4.3. ProMES design criteria: design process, design, and implementation.

The ProMES method has some of the design criteria for productivity measurement systems 'automatically' built in because the method 'per definition' generates a system which provides an overall index of productivity and weighted subindices of productivity on a common metric (see Section 2.1.4). Whether the other design criteria mentioned (e.g. accurate reflection of organizational policy, flexibility, acceptance) are met depends on the specific implementation of the method. Latter criteria will therefore be included in the normative control loop model and discussed in this section.

ProMES design-process criteria

Three design process characteristics of the ProMES method represent an essential precondition for a *high-quality system* which is *accepted by personnel and management* (Pritchard, 1990). These characteristics are the participative design procedure, review and approval by management, and discussion until consensus.

Participative design. Involving the employees who are supposed to use the system in its development is considered essential to employee acceptance of the system *and* the quality of the system. Systems imposed without input from unit members will probably be less accurate, since unit personnel are the experts on the work they do. In some areas of the work their expertise will exceed that of their supervisors and the management. The perceived validity and accuracy of an imposed system will also be low, because of a feeling that the program does not fit the unique needs and context of the unit. Furthermore, unit personnel will resent the implication that they–the experts on their job–have nothing to contribute to such a program. The participative approach creates 'ownership' of the resulting system. It also promotes acceptance because of a higher degree of understanding of the system by unit members.

So, having unit personnel participate in the system's development will increase acceptance, ownership, and understanding of the system and reduce the likelihood that the system is ignored or sabotaged by the unit. Additionally, using the expert knowledge of the unit personnel will increase the validity and accuracy of the system.

Review and approval by management. The participative, bottom-up approach taken in developing the ProMES system produces a system which measures what should be measured and which is accepted by its future users. Review and approval meetings in which the preliminary system is discussed with management produce additional benefits. Firstly, unit personnel must defend the system to management. This enhances their sense of ownership of the system. Secondly, management has the opportunity to review the system and suggest modifications which strengthen the link between management policies and the elements and priorities within the measurement. This enhances the perceived validity of the system from the management's point of view and thereby management commitment to the system. Finally, the process of discussing and working out differences of opinion between management and unit personnel increases communication between levels in the organization and may, in some cases, constitute the first real dialogue between management and personnel.

Discussion until consensus. The basic process to be used in developing a ProMES system should be discussion until consensus is reached. The design team discusses the issue at hand until there is general agreement on the solution. On the major issues unanimity should be obtained. A strong majority consensus should be reached on the other issues, both through interactive discussion. This interactive discussion, in which all group members should have the opportunity to participate fully, is considered essential to the success of the method. Different individuals in the unit will have different ideas on the responsibilities of the unit, which measures should be used, etc. These different ideas are typically a result of looking at the work form different perspectives, and usually all have some validity. Exposing all the members of the design team to these different views broadens the perspective of all the members and results in a better system (hence: 'constructive disagreement is good'). Decision-making techniques that decrease face-to-face discussion are therefore advised against.

Another factor which should be considered is the *organizational resources* that are needed to design ProMES. A number of parties have to attend design meetings (group members, supervisors, facilitators, and–on some occasions–management). In addition, some tasks will have to be performed between meetings. All parties concerned should be willing to commit to these time-investments.

ProMES design criteria

The ProMES system which results from the design process should possess desirable characteristics with regard to the validity of the system, the measurement process, and the way the system deals with interdependencies.

Validity of the system. The four main elements of the system (products, indicators, contingencies, and feedback report) should each meet a number of criteria (Pritchard, 1990; Schoonen, 1993).

Products: 1) the unit is responsible for the products; 2) the products are clearly stated; 3) if the unit does exactly what the products say, the organization benefits; 4) the products are general; 5) the set of products is complete.

Indicators: 1) the indicators cover all products; 2) the indicators cover each product completely; 3) the indicators are valid (i.e. an accurate index of product accomplishment); 4) the indicators are controllable by the unit; 5) it is cost effective to collect the indicator data; 6) the indicators are understandable to personnel of the unit.

Contingencies: 1) the maximum on a contingency is the maximum feasible value that the unit can achieve on an indicator; 2) performance at the zero point of a contingency is neither good nor bad; 3) the minimum on a contingency is the point at which negative consequences would start to happen if the indicator got that bad; 4) the relative importance of the indicators is accurately reflected.

Feedback report. The following elements should be included in the feedback report: 1) a list of products and indicators and the level of each indicator for the period, along with its associated effectiveness value; 2) the overall effectiveness score (possibly also expressed as a percentage of maximum effectiveness); 3) historical data and the amount of change between the historical data and the current data; 4) historical data over a long period of time (presented in a graphic form); 5) a clear set of priorities for improving productivity; 6) a comparison of percent of maximum with other units (if applicable).

Process of measurement. The process of performance measurement should be feasible, it should not be not too elaborate and complex. If productivity is too elaborate or if it takes to much time, it might not be executed. A good measurement system that is not used because collecting the necessary data takes too much efforts has not reached its goal. This has to be taken into account when developing the system. Relevant aspects are (Schoonen, 1993): the number of indicators, the complexity of measuring the indicators, the complexity of combining the scores, and the complexity of combining feedback reports.

Interdependencies. For the members of a group to be able to identify with the group's productivity, it is important that the group be not too large. For controllability reasons, it is important that no groups that are heavily dependent on each other for their productivity be separated. The appropriate group size is therefore the smallest unit that does not separate groups that must work interdependently (Pritchard, 1990). The role of task interdependence was discussed more extensively in Section 2.2.5.

ProMES implementation criteria

In addition to the general criteria for effective feedback and goal setting, a number of criteria specifically related to the feedback meetings need to be mentioned. Also, in this section, the desired consistency between goal setting and feedback systems and other control systems in the organization is discussed.

Feedback meetings. The following guidelines for feedback meetings are supplied by Pritchard (1990):

- a meeting is held between supervisor and unit personnel after each feedback period to discuss the feedback report;
- incumbents are willing to invest time in feedback meetings where ideas for productivity enhancement are created;
- areas of productivity increase and decrease are explored in the feedback meetings; needed changes are identified, implementation strategies are discussed, and plans are made for the next feedback period;
- the feedback meeting is conducted as a fact-finding exercise, not a search for excuses or a method for attaching blame;
- if feedback pertains to individuals, it is given privately; only the individual and his/her supervisor are given the report; if feedback pertains to a group, it may be public.

Flexibility. Organizational goals and policies may change over time, and objectives, tasks, and priorities of the units working with ProMES may change accordingly. The ProMES must be able to incorporate such changes without much difficulty. For example, priority changes may be incorporated in the system through a re-evaluation of the relative importance of existing indicators and corresponding changes in the contingencies. Major changes in work procedures, equipment or number of personnel may also require revision of contingencies. The addition of a new task may require the addition of one or more new indicators or even a new product (the reverse is the case if a task is discarded). Pritchard (1990) suggests a periodic review of the system (at least once per year) and an additional review of the system after any major change in the unit's tasks, equipment, or personnel.

Consistency with other control systems: the performance appraisal system. According to Ilgen et al. (1979), the willingness to respond to feedback is influenced by the extent to which the feedback recipient believes the feedback source influences the contingency between the recipients behavior and his or her receipt of valued outcomes. Thus, the more valued outcomes are associated with the feedback, the higher the perceived importance of feedback will be, and the lower the likelihood that the feedback will be ignored. When valued outcomes are mainly associated with issues which are not covered by the feedback system, dysfunctional effects may occur (Connellan, 1978). A control system which is often associated with a feedback (and goal setting) system is the performance appraisal and rewards system. An example, based on Algera (1990), illustrates the possible dysfunctional effects of inconsistency between feedback systems and reward systems: rewarding only sales volume and not after sales services can lead salespeople to pay attention only to sales volume, with possible dramatic negative effects for long term sales because of neglecting after sales services and thereby losing customers in the long run (even though management insists both short term and long term sales are important!). This demonstrates the importance of one of the reinforcement principles: the reinforcer must be made contingent upon the (all!) desired behaviors.

In most cases, the ProMES method will result in a system which measures group productivity. According to Pritchard (1990), the issue of consistency of the ProMES system and individual performance appraisal and rewards systems for group members will, in most cases, have little relevance, since individual contributions to the productivity of the group are difficult to separate because of intragroup dependencies. However, if the productivity measurement system is based on individuals working independently, consistency between both systems should be pursued.

Irrespective of the level of measurement (individual or group), the consistency between ProMES and the performance appraisal system of the unit supervisor must be considered, since unit productivity can be an important element in the performance appraisal of the unit supervisor.

2.4.4. Organizational context: situational constraints

Pritchard (1990) and Schoonen (1993) list a number of organizational characteristics that should be present to increase the probability of a successful design and implementation of a ProMES system. These characteristics refer to the initial state of the organization and the initial attitude of the organization towards productivity and ProMES. From both categories, a selection of characteristics (relevant for this study) will be described in some detail.

It should be noted in advance that very few organizations will completely conform to these criteria (and those that do might have less need for a productivity enhancement system than those that do not). Also, the relative importance of the criteria is unclear and may depend on other characteristics of the organization. Moreover, it will not be completely clear at the start of a project how an organization 'rates' on each of the criteria (unless an accurate assessment can be given through an extensive 'feasibility study'). So, most or all ProMES projects will start in a less-than-perfect organizational context, the intricacies of which are not completely clear to the researcher/consultant. As the organization possesses fewer of the desired characteristics, the design problems encountered will be more difficult, and the probability of failure will increase. The effectiveness of the ProMES control loop will depend to a large extent on the implementation of satisfactory solutions for specific situational constraints.

Initial state of the organization

The presence of mutual trust and respect between incumbents, the supervisor, and management. A productivity measurement system such as ProMES requires that unit members, supervisors, and higher management work together in an atmosphere of mutual trust and respect (Pritchard, 1990). If employees feel that they would be making a system that makes them accountable to a management that may use the system against their interests, they may be unwilling to participate in the development of a productivity measurement system or sabotage it if such a system is imposed on them. So, as a necessary precondition for conducting a successful ProMES program, employees should at least have a minimum amount of trust in their management. On the other hand, management must at least have some level of trust and respect for the employees, since the bottom-up development process requires management to openly discuss the contents of the system with unit personnel on a more or less equal basis.

Provided a minimum 'threshold' amount of mutual trust is present, the participative design process, which requires horizontal and vertical communication may help to enhance the degree of mutual understanding.

The implementation of the ProMES system by means of feedback meetings may also be affected by the issue of trust. For example, Nathan, Mohrman jr. & Milliman (1991) found that the ongoing interpersonal relationship between supervisor and subordinate influenced the effectiveness of performance appraisal interviews. This may be a relevant finding in case a ProMES system were in some way linked to individual performance appraisal of subordinates.

The stability of the organization management. If turnover is expected among key management personnel during system development or after system implementation, an effective ProMES system will be difficult to develop and maintain, since policies and priorities may change frequently due to possibly different views of the new management on issues related to performance management. Pritchard (1990) therefore suggests to wait for expected changes to take place before starting development and making sure of top management support, so that eventual changes at lower management levels will pose less of a problem.

The familiarity of the organization with productivity measurement (and ProMES). If an organization has had a lot of experience with measurement, the introduction of a ProMES system may be a natural extension of measurement already done. The organization will be aware of problems that may occur and will be able to anticipate them. If measurement is a rather foreign concept to an organization, developing and implementing a system such as ProMES would be a new experience, requiring education on measurement issues and more support from outside specialists.

Specifically, the existence of successful ProMES systems in other parts of the organization may further the enthusiasm of management for another project and stimulate other units to start a ProMES development process (Schoonen, 1993).

The possibility of substantial productivity improvement. The willingness of management to support a ProMES program will depend on the extent to which management feels that substantial productivity improvement is possible, and programs aimed at increasing productivity through motivation have not been installed (or have been installed ineffectively). If, for example, previous efforts aimed at increasing productivity through motivating employees have already resulted in substantial productivity improvements, management may not consider it cost-effective to conduct a ProMES program.

Amount of vertical job specialization. Vertical job specialization separates the performance of the work from its administration (Mintzberg, 1979). It influences the amount of control a unit has over its task performance (e.g. 'ability to intervene' in the control loop model). This determines to a large extent the features of the productivity measurement system, since controllability is an important aspect of indicators and goals. The less control the unit has over its performance, the narrower the focus of the productivity measurement system will be (Schoonen, 1993).

Initial attitude of the organization towards productivity and ProMES[6]

Organizational attitudes: productivity improvement is considered important, not easy, and a long range organizational change effort. If the organization is doing very well, it may not consider productivity improvement a priority. This complacent attitude may lead to problems in the future, and it makes the institution of a productivity measurement and improvement system difficult. So, productivity improvement should be considered important and required (Pritchard, 1990). Secondly, the effort that should be given to improve productivity should not be underestimated. As Pritchard (1990) states: "Productivity improvement is thus something that takes work and constant vigilance in that we all like routines and once a set of procedures is learned, there is considerable resistance to change. It takes commitment and dedication to break out of the old patterns and to systematically develop a productivity measurement system and to use it to make changes". Lastly, productivity improvement should not be seen as a separate peripheral activity that will last a fixed time period (until another project comes along). Successful productivity improvement programs are long range efforts that are permanent parts of the organization's operations. "No productivity improvement program will provide an overnight answer to problems that may have been years in the making" (Pritchard, 1990).

Management shows commitment to the productivity enhancement program. It is critical that management commits to the productivity improvement program. Management should consider the program to be important, commit resources to it, secure adequate and ongoing support, be interested in its results, and protect the program when it is challenged. This management commitment should be *visible* to everyone in the organization. It is imperative that the most senior manager in the functional area where the project will be done is committed to the project (Pritchard, 1990). The importance of management commitment is illustrated by a meta-analysis study by Rodgers & Hunter (1991) of 70 Management By Objectives studies (defined as a combination of goal setting, participation in decision making and objective feedback; the meta-analysis

[6] A number of desirable organizational characteristics refer to both the objective and the perceived state of the organization. Both requirements should be fulfilled. For example, organizational productivity should not only be largely dependent on the effort of the personnel subsystem, management should also believe this to be the case. To avoid redundancy in the description, such desirable characteristics will feature in one of the two categories.

included the ProMES studies by Pritchard et al. (1987)). The results showed that when top management commitment was high, the average gain in productivity was 56%. When commitment was low, the average gain in productivity was only 6%. Furthermore, a survey among 236 top-level executives in the United States revealed that top management commitment and involvement was considered one of the main factors influencing success or failure of productivity improvement efforts (Judson, 1982).

Management belief: organizational performance depends heavily on the effort of its human resources. ProMES is an elaboration of the motivational approach toward improving organizational productivity (see Section 2.1.3). The more organizational productivity depends on the efforts of its human resources, the more management will be willing to commit resources to a program aimed at improving the personnel's productivity.

This general overview of situational constraints concludes the theoretical basis for the study. In the next chapter, which provides a broad outline of the study, some specific characteristics of the research setting are added to the above overview.

Chapter 3

Outline of the study

In this chapter, we present a broad outline of the study. The research objectives, the resulting research questions, and the research design are discussed. The research setting–the field service department of Nashuatec–is described in some detail. Also, the framework of the 'design cycle' is applied to the development of a ProMES system. In the last part of the chapter, the conditions that are distinguished in the research project are described in terms of one of the theoretical concepts: participation.

3.1. RESEARCH OBJECTIVES

The first objective of the study is to determine whether the ProMES method for designing performance management systems can be successful in a context which differs in several respects from the context in which the method was tested for the first time.

The study consists of two phases: first, a ProMES system is designed and implemented in two regions of a field service department. Subsequently, the design is transported (i.e. implemented without going through the participative design process) to six other regions. Each of these two phases contributes to answering one of Pritchard et al.'s (1989) research questions (summarized in section 1.4.3).

The generalizability of the ProMES approach that was very successful in the US Air Force setting is tested by applying it to a very dissimilar setting: the field service department of a Dutch supplier of office equipment, named Nashuatec. The settings differ in several respects (see Table 3.1). First, the Air Force organization is an American government organization, while Nashuatec is a private Dutch/European service organization. Second, whereas the Air Force base represents a local organization structure, Nashuatec has a geographically dispersed organization. Third, whereas the Air Force units for which the ProMES system was developed performed group tasks, the field service units consists of technicians primarily doing an individual job.

Table 3.1. Comparison of 'US Air Force base' and 'Nashuatec'

Characteristic	US Air Force base	Nashuatec
Country	USA	The Netherlands
Industry	Government (non-profit)	Private (for profit)
Organization structure	Local	Geographically dispersed
Type of units (number)	maintenance (1), supply (4)	Service/maintenance (14)
Type of task	Group task	Individual task (some interdependencies)
Reward issues	Group incentives	Individual performance appraisal

Related to the latter difference is the issue of individual performance appraisal and rewards, which is mainly relevant in case of individual productivity measurement; this represents the fourth difference between the two settings. So, it remained to be seen whether the ProMES approach could be successfully applied in the Nashuatec setting, and what elements would have to be changed or added to the approach to accomplish this.

Testing the transportability of the system is the secondary objective of the study. In a lot of organizations, units exist that have essentially the same function but that are separated by time or distance (e.g. a chain of grocery stores, regional sales offices of a large firm, units working on a shift system). For these types of organizations, the role of a participative development procedure in obtaining an accepted control loop determines the approach that has to be taken and its corresponding cost-effectiveness. For example, if it were possible to participatively design the system in one organizational unit and then transport the system to all comparable units within the organization with similar success (acceptance of the system, productivity gains), a large-scale institution of ProMES control loops could be achieved at relatively low cost. If, on the other hand, participative design turns out to be a necessary precondition for success in that the participative design process has to be carried out in each comparable unit, instituting a ProMES system would be much more costly.

From a scientific point of view, the transportability issue is also an interesting one. The participation literature, reviewed in Section 2.3, basically reports mixed results on the relationship between participation in decision-making and productivity. On the other hand, a participative development procedure is considered an essential prerequisite for a successful ProMES program (see Section 2.4.3). A comparison of the effects of participative design and implementation of a system to the effects of non-participative transportation of a system could give an indication as to the importance of participation in the design of ProMES systems.

The organizational structure of Nashuatec's field service department–several comparable, geographically dispersed regions–provided an excellent opportunity to assess the effects of transporting the system. It was decided to develop the system participatively in two units and transport it to the remaining units, i.e. implement it without going through the participative development process.

3.2. DESCRIPTION OF THE SETTING

This section supplies some background information on the organization in which the project took place, the job of a service technician which was the focus of the ProMES system, reasons for stating the project, and situational characteristics which determined to a large extent the specific design problems that had to be solved.

3.2.1. The organization

Nashuatec[1] is a large Dutch supplier of office equipment (photocopiers, fax machines and laser printers). The company does not manufacture these machines, but buys them from leading Japanese suppliers, sells them under its own name and provides service on these machines. The company's major departments are sales and service. In all, about 650 people are employed at Nashuatec. The company's customers range from very small (e.g. private persons, small companies) to very large (e.g. multinationals, universities, ministries, and municipalities).

Nashuatec offers a large assortment of photocopiers, ranging from small, inexpensive copiers for infrequent use ('low volume' copiers) to large copiers with several additional features for intensive use ('high volume' copiers). Although this current assortment usually consists of about 15 different types, the company still provides service on approximately 75 types of copiers sold in the past.

The service organization

Since photocopiers (and to a lesser extent, fax machines and laser printers) require regular maintenance, service is an important part of Nashuatec's operations. The organization chart of Nashuatec's service department is shown in Figure 3.1. The head of the service department (a member of the company's management team) is the primary supporter of the ProMES program, together with the personnel manager. The largest unit within the service department is the field service department. We will focus on this department, since the ProMES program took place in its field service regions.

At the head of the field service department are two field service managers, responsible for the service reception unit, the service planning unit, and the field service regions. The field organization is organized geographically and consists of 14[2] regions. Each region consists of 20 to 23 technicians and a supervisor.

Most technicians have a polytechnic education in electro-technics. The majority of technicians is between 25 and 30 years old. Depending on ability and experience, the technicians have a package of 2 to 10 types of photocopiers[3], for which they possess

---

[1] Until 1990, the company was part of the office systems division of the Nashua corporation (USA). The company then became a part of Gestetner Holdings (UK) and was briefly known as Gestetner/ Nashua. In November 1992, the company changed its name to 'Nashuatec' (its juridical name is 'NRG Nederland BV'). Nashuatec's head office is located in 's Hertogenbosch, the Netherlands.

[2] Due to a number of factors, the number and size of regions to which the technicians were allocated changed a few times during the period this thesis reports on. The number of regions varied from 11 to 14. The number of regions which existed for the longest period of time during the research period was 14; this will be used throughout this thesis.

[3] The vast majority of service resources were spent on repairing and maintaining *photocopiers*: more than 90 percent of all calls (visits for repair and maintenance) pertained to photocopiers. As

the specific knowledge needed to carry out repairs and maintenance. About 10 percent of the technicians (two per region) are so-called senior technicians. They are specialists in certain types of copiers. Each technician has his[4] own company car and a car stock of spare parts which are needed frequently.

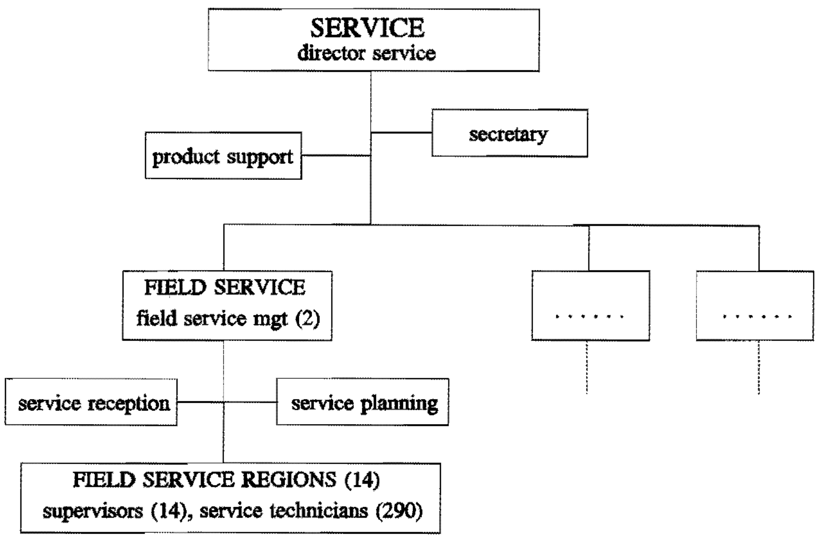


Figure 3.1. The service department of Nashuatec.

3.2.2. A technician's job

A technician's job typically consists of repairing machine malfunctions and carrying out preventive maintenance. The maintenance is done at the same time as the repairs, which means that there is no separate schedule for preventive maintenance. The regular course of action is as follows: the client makes a phone call to the service reception, indicating that his photocopier is out of order. A first attempt is made by the service

will be seen in Chapter 4, the ProMES system at Nashuatec was designed for the part of the technicians' job involving photocopiers. It was decided to leave fax machines and laser printers out of the ProMES to be designed, due to technical and procedural characteristics which would probably necessitate a different measurement system. The description at hand therefore focuses on photocopiers.

[4] During the four year time-period this thesis reports on, only one female service technician was employed at Nashuatec. While acknowledging her presence, for simplicity's sake (and because this –unfortunately–still is an accurate description of reality), male pronouns will be used when referring to service technicians.

reception employee to solve the problem by giving the client some instructions on the phone. If this does not succeed, the problem is transferred to the planning department, which instructs a technician to visit the client, repair the machine and carry out the preventive maintenance suitable at that time, given the maintenance history of the machine. The technician drives to the client's location, repairs the machine, follows the prescriptions with regard to preventive maintenance, fills in a card containing information on the machine's repair history, and leaves for the next client. Before he does so, the technician enters data by phone into the company's information system, regarding, among other things, the time it took to travel to the customer, the time needed to repair and maintain the photo copier, the spare parts he used, and the number of copies on the counter. He also orders additional spare parts to replenish his car stock (the supplies asked for are delivered once a week at the technician's home address by a special service, or the following night by a special courier, if necessary). The technician then contacts his planner in order to receive his next job. It is important to know that his next job can be with any client in his region: the planning department allocates technicians to jobs in such a way that the available technicians are used efficiently and agreements with clients regarding service within a certain time limit are fulfilled. For that reason, there are no fixed connections between technicians and clients within a region: any technician can be allotted a particular repair, provided that he has the requisite knowledge. This causes interdependence between technicians, which is discussed in Section 3.2.4. Apart from their colleagues, technicians are dependent on a number of other groups inside and outside the organization. Three examples will be given. The timeliness with which the warehouse fulfills requests for spare parts influences the technician's stock of parts he carries in his car, and in turn, the percentage of repairs that cannot be finished adequately because of a lack of spare parts. In such cases a return visit is necessary (the following day). Although a technician is primarily dependent on his own knowledge and skill, he can ask for help from a product support specialist over the telephone, if he is unable to solve a technical repair problem on his own. If that does not solve the problem, a senior technician may be called in. Lastly, the technician is dependent on the client. The way a photocopier is handled by the client influences the occurrence of malfunctions (e.g. a photocopier for general use in a university corridor or a copy-shop may require more frequent repairs and maintenance than a copier which is used exclusively by the secretary of the board of a law firm).

3.2.3. Management's reasons for starting the program

There were three main reasons for the management at Nashuatec to start a ProMES program. First, ProMES could help maintain or, if possible, improve the quality of service while reducing its cost. Since the 1970s, the office automation industry had moved from a product-oriented industry via a price-oriented industry to an industry in which companies that were able to meet the high demands placed on service had the competitive edge in a situation where product features and prices hardly differentiated among suppliers of photocopiers. Nashuatec had been a leading company in this respect, e.g. through the use of effective and efficient maintenance procedures and sophisticated planning systems. In order to keep this competitive edge, attention for a high-quality, cost-effective service performance was considered of vital importance. ProMES could help direct the technicians' performance towards this objective.

A second reason for starting a ProMES project originated from the service technicians' recurring requests for valid feedback about their performance. A questionnaire measuring the extent to which feedback and goal setting mechanisms operated effectively in the field service regions, filled out by 133 service technicians as part of preparatory study (Van der Putten, 1988), showed that a number of criteria for effective feedback and goal setting were not fulfilled. Although a large amount of data concerning the performance of technicians were collected, there was no regular feedback. Data that were fed back were often inaccurate, difficult to interpret, too late, and/or involved uncontrollable dimensions. Most technicians expressed a need for meaningful feedback about their task performance. The ProMES approach was considered a promising method for designing a valid measurement system that could provide the valid feedback asked for.

A final reason involved the possible future use of ProMES data for performance appraisal purposes. Due to the highly isolated nature of the work of service technicians, providing valid information that could be used in a performance appraisal had always been difficult, especially since quantitative performance measurement had yielded highly debatable input for performance appraisal. Management hoped that a productivity measurement system designed according to the ProMES method would generate accepted input for performance appraisal. This would necessitate a measurement system which measures the productivity of *individual* technicians and which could *uniformly* be applied to all service technicians.

3.2.4. Main characteristics of the setting

Some of the company's general characteristics are quite relevant for understanding the design problems run into during the process of system design and implementation. The five main characteristics will be discussed.

Interdependence between technicians within an individual task

On average, about 50 percent of the client calls a technician carries out involve a photocopier the same technician has visited on the previous occasion. So, a technician is often confronted with the results of repairs and maintenance carried out by a colleague of his at an earlier date (which may have been several months in the past). If that colleague has carried out maintenance according to specifications and has provided relevant information on the activities he carried out (e.g. which spare parts did he replace, what type of maintenance did he carry out?), the next technician working on the machine will likely have a sound basis for starting his work on the machine. On the other hand, if a technician has not carried out the required maintenance and/or supply relevant information, the next technician will encounter a number of avoidable problems. For instance, he will have to spend more time to carry out back maintenance or he might need more time for fault diagnosis.

With the management's desire to eventually use ProMES information in the technician's individual performance appraisal in mind–which would necessitate an individual measurement system–this interdependence is likely to become a complicating factor. Yet, solely designing a group system (per region) will in all probability diminish its usefulness, not only because the system cannot, in the long run, contribute to individual performance appraisal of technicians, but also because individual accountability for what is in essence an individual task will be lost (see Section 2.2.5).

Complexity of work flow

Because Nashuatec provides service on approximately 75 types of copiers, each technician specializes in a subset of this population. Technicians who are new at the company have received training on one or two relatively simple types. Experienced technicians specialize in five to ten different types. The higher the structural need for service on a type of copier–which depends on the number of machines 'in the field' and the malfunction frequency–the more technicians have knowledge of that particular type and visited it regularly.

It is known from earlier attempts to measure the productivity of technicians, that technical characteristics of the copier determine to a large extent which results are obtained. A quality measure called 'Mean copies between calls' (the average number of copies a client can make between two consecutive calls, abbreviated: MCBC) is a clear example of this. On a small type of copier, the MCBC may average 10,000, whereas it may average up to 60,000 on large types. A comparable situation occurs with a number of other measures management has been using for feedback and performance appraisal purposes. This poses a virtually unsolvable measurement problem, since it is not possible to directly compare the MCBC results of technicians who work on different types of copiers. Furthermore, comparing a technician's productivity from one period to another is difficult, since the machines allotted by the planning department to a technician vary from one period to another.

Lack of horizontal and vertical communication

The geographical structure of the service department has caused a general lack of communication, both *horizontally* (within regions) *and vertically* (between regions and service management).

As mentioned before, a technician's job is in essence an individual one. The technician works on his own and only meets his colleagues at monthly meetings of the region and at infrequent technical meetings. At most once per month, he is visited by his supervisor when working at a client's office. Consequently, technicians have little opportunity to communicate with each other and with their supervisor.

Almost all communication between the regions and the management at the company head office occurs through memorandums or the region supervisor, who visits the head office a few times per month. Visits from field service managers to the regions are infrequent. Most technicians have been at the head office only once or twice, when applying for the function of service technician. Because of this lack of communication, all kinds of prejudices and distrust easily arose and have continued to exist.

'Top down' organizational culture & cultural differences

The company's culture at the start of the ProMES project may be characterized as 'top down': decisions are made at the head office; technicians should do as they are told. This attitude has led to irritation in the field service department and some long-running disagreements. Also, projects initiated by management are usually greeted less than enthusiastically in the field service department and are often terminated for lack of success.

In addition, large cultural differences exist between the sales and the service department, and between the head office and the field: the sales department and the head office could be characterized as a white collar culture, field service as a blue collar culture. In addition to that, some feelings of inequity exist in the field service regions because sales representatives and managers can earn significant incentives, whereas service technicians cannot.

The combination of the lack of communication and the top down culture constitutes a less than optimal starting point for the ProMES program. Because of the limited possibilities for interaction within a region, the organization of ProMES development meetings will not be easy. The very *limited amount of mutual trust and respect* (especially between regions and management) is a clear deviation from one of the 'initial state of the organization'- criteria (Section 2.4.4).

Performance appraisal and rewards issues

There seems to be a dilemma concerning the use of ProMES information in the performance appraisal of individual technicians. If the ProMES system is not going to be used for performance appraisal purposes, the design process will probably be easier (e.g. less fear of being exploited, easier acceptance of indicators which are not a hundred percent controllable). Yet, it is doubtful whether productivity improvement activities which are not consistently linked to performance appraisal will produce the desired effects. To rule out any use of individual ProMES data in the performance appraisal–a supervisor who has the ProMES information pertaining to individual technicians at his disposal may be tempted to use it anyway–only group productivity should be measured. This would, however, virtually preclude any improvements, due to non-existent individual feedback and accountability (see Section 2.2.5). If, on the other hand, one of the objectives of the ProMES program were the use of the productivity data as input for performance appraisal, the design process may be very laborious, with a higher likelihood of the system being rejected or manipulated in an adverse context such as this setting. If, however, a valid, accurate, and accepted system were to result from the design process, an essential prior condition to success–consistency between

ProMES and appraisal and reward systems–would be fulfilled. This would increase the probability of productivity improvement.

Concluding remarks

While all context factors listed in Section 2.4.4 have some relevance for the process of designing and implementing a ProMES system in the Nashuatec setting, the design problem is probably best illustrated by the five situational characteristics mentioned above. This specific combination of constraints suggests that designing and implementing an effective ProMES control loop in the Nashuatec setting will be a difficult undertaking in which unique solutions will have to be developed for unique design problems. Table 3.2 summarizes the objectives the ProMES system should fulfill and the main problems that should be solved during the design and implementation of the system.

Table 3.2. Objectives and design problems.

Objectives	Design problems
1. Improvement of productivity of service technicians 2. a. Accurate measurement of productivity of individual technicians and service regions 2. b. Accurate feedback 3. Contribution to more accurate performance appraisal (long-term)	1. Interdependence between technicians within an individual task 2. Complexity of work flow 3. Lack of horizontal and vertical communication 4. Top down organizational culture & cultural differences (3 & 4 contribute to lack of mutual trust/respect) 5. Performance appraisal and rewards issues

3.3. PLACING THE PROMES APPROACH WITHIN A DESIGN CYCLE FRAMEWORK

3.3.1. Introduction

In this study, the ProMES approach is placed within a design framework. In this design framework, the ProMES approach is viewed as a method for designing a product which fulfills one or more functions (e.g. measure productivity, feed back productivity information, provide data for performance appraisal). A 'list of requirements' is the basis for the design process and the product (e.g. high employee involvement, valid contingencies). Characteristics of the provisional design are compared to the requirements. If this prototype does not conform to the requirements, an iterative process occurs in which previous phases in the design cycle are repeated, until an acceptable design results. Interventions in the design and implementation processes are guided by the accepted control loop model and the criteria for an effective control loop, as set out in Chapter 2.

In this study, it is not our purpose to test the validity of the design criteria and situational constraints in the 'accepted control loop'-model. Whereas evidence for some elements of the model, such as the general design criteria for effective feedback and goals, has been very well established, other elements of the model are of a more tentative nature. For example, the list of context factors is not exhaustive and the selection made for the Nashuatec context is a subjective one, based on information obtained before the start of the design process (literature on ProMES, questionnaire data, observations, task descriptions, etc.) and on experiences during the early stages of the design process. Analysis of ProMES projects carried out in different settings could contribute to improving the tentative parts of the model. In this thesis, the model is used to aid in assessing the effectiveness of the interventions made in the design and implementation processes.

3.3.2. Design of a ProMES system: steps of the design cycle

In this section, the ProMES design process at Nashuatec is described in terms of the design cycle (discussed in general terms in Section 1.5.2 and shown again in Figure 3.2).

Analysis. The ProMES system, to be designed in the specific context of the field service organization, should possess the following *functions*. The system should measure the productivity of individual technicians and groups of technicians in a valid way. It should improve their productivity through a feedback procedure. Furthermore, the ProMES productivity data should in the long run be usable as a part of the performance appraisal of individual technicians.

The *list of requirements/criteria* is based on a number of sources. The literature on feedback and goal setting (e.g. Ilgen et al., 1979; Locke & Latham, 1990) presents a number of general criteria that should be met by a control loop for performance management. Added to these are the criteria posed to the productivity measurement system which generates the necessary feedback information (e.g. Pritchard, 1990). There are a number of characteristics that the design process should possess (e.g. Pritchard, 1990).

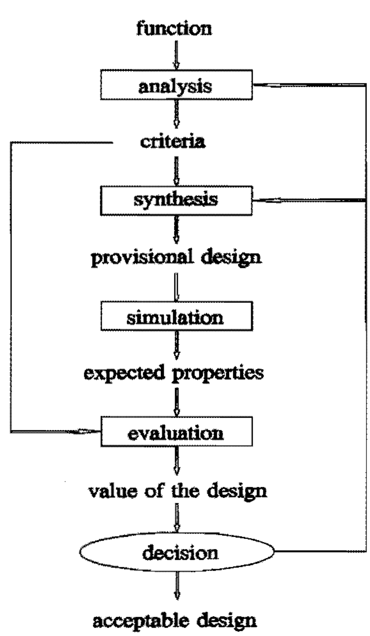


Figure 3.2. The design cycle (Roozenburg & Eekels, 1991).

Finally, the context in which the system should operate produces a number of specific problems. These are translated into design criteria.

The criteria that are employed for the design process and the constraints that are posed by the context in which the control loop should operate are central to this thesis, in that they pose the essential research problem: in the given context, it will be very difficult to design a ProMES control loop in full accordance with the list of requirements.

Synthesis. In designing a ProMES control loop, the *design process* followed is crucial. Therefore, the starting point of the synthesis step is a design procedure which should maximize the possibility of achieving a high quality design accepted by the employees who should use it to control their performance. A design procedure is chosen that meets the demands in the list of requirements as closely as possible, taking into account constraints posed by the context. The ProMES control loop is developed participatively by a design team which incorporated the employees from two regions who would be the first ones to work with the system. The synthesis step results in a *provisional design* of the ProMES system 'on paper', that appears to meet most of the demands posed in the list of requirements.

Simulation. The simulation step in the design of the ProMES system differs from the contents of the step as given by Roozenburg & Eekels (1991). Roozenburg and Eekels view simulation primarily as a mental exercise, which is supported by some kind of abstract model of the product-to-be. The actual production does not take place until the whole design cycle has been finished. For this purpose, two extra steps are added to the cycle: realization and testing of the product. In designing the ProMES control loop, the simulation step is to a large extent an empirical one. It involves an implementation period of a 'prototype' in two groups that took part in its design.

Evaluation. The evaluation step involves a 'participative problem analysis', following the simulation period in which feedback was provided by the 'prototype'. The design team evaluates the properties of the design and generates ideas for improving the design and/or improving the extent to which situational constraints have been met. The normative 'accepted control loop'-model is used as a reference point for choosing the appropriate solutions from the alternatives that are generated. In an iteration, in which the synthesis step is redone, necessary adjustments are made to the design. The improved design-prototype is then re-implemented in the two regions that participated in the design and evaluation.

Decision. Generally, this step in the design cycle involves a decision on whether the design can be taken 'into production' or a continuation of the process of re-designing is necessary in order to generate a better design. In the context of this study, 'taking the design into production' refers to the transportation of the system to the remaining service regions. This transportation process could be considered a separate design process, in which criteria regarding acceptance of the design and the time needed to implement it become more relevant than criteria regarding quality of the design. In transporting the system to other regions, the quality of the design (achieved through the participative design process) is not compromised. However, the participative design process which should contribute to the acceptance of the system by its future users is absent, mainly due to time considerations. This might result in lower acceptance of the system in the transport regions and, consequently, in lower effectiveness of the control loop. This is a central issue in this thesis.

3.4. RESEARCH QUESTIONS AND RESEARCH DESIGN

The two main research questions correspond with the issues of the generalizability of the ProMES method and the transportability of a ProMES system.

3.4.1. First phase of the study: generalizability of the ProMES method

The Nashuatec context appears to be an especially difficult context for designing an effective control loop. In Section 3.2.4, five situational characteristics were identified that should be taken into account when designing a ProMES system: 1) individual task with interdependencies, 2) complex work flow, 3) lack of communication, 4) top-down culture/lack of trust, and 5) performance appraisal and reward issues. These characteristics suggest that a number of issues will have to be resolved in order to design an effective ProMES control loop. This leads to the following research question:

"How can an effective control loop be designed in the Nashuatec setting with its specific characteristics?"

The first phase of the study involves the participative design and the implementation of a ProMES system in two regions of Nashuatec's field service department. A brief outline of this 'participation' phase and the activities involved is given in Figure 3.3 and Table 3.3. The duration of this first phase of the study will depend heavily on the progress made in the design and implementation processes (i.e. the number of iterations that have to be made before a final, acceptable design is realized).

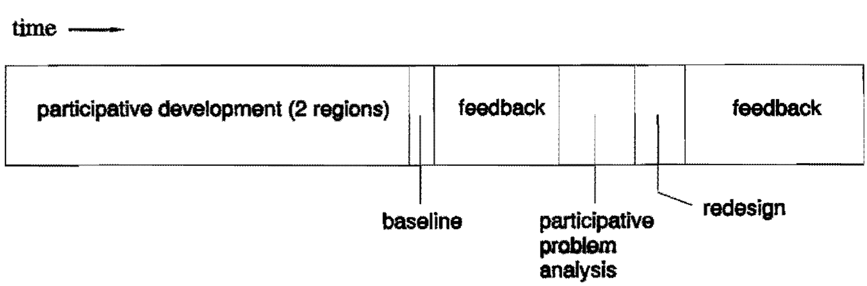


Figure 3.3. The participation phase: a chronological overview.

Table 3.3. The participation phase: overview of activities.

Activities	Contents
Participative design	Two regions participate in the design of a ProMES control loop (in two parallel design processes).
Baseline/technical implementation	The program which generates the ProMES information is completed. This period is used as a baseline period.
Feedback	The 'prototype' ProMES system is implemented by means of regular group feedback meetings in both regions.
Participative problem analysis	Both design regions participate in an evaluation of the 'prototype' system and the feedback meetings. Ideas are generated for improving the design or removing situational constraints. Desired adjustments are discussed with the management. Feedback continues during this period.
Redesign	The adjustments agreed upon are implemented into the design. No feedback is provided during this period.
Feedback	The final design is implemented.

A quasi-experimental design is used to test the effects of the implementation of the system. The design resembles the 'interrupted time series with a nonequivalent control group time series design' (Cook & Campbell, 1979):

experimental condition $O_1\ O_2\ ...\ O_n\quad X\quad O_{n+1}\ O_{n+2}...O_{n+m}$

control condition $O_1\ O_2\ ...\ O_n\qquad O_{n+1}\ O_{n+2}...O_{n+m}$

In this design, the experimental treatment (X) is given in the experimental condition in which multiple measurements (O) are done before and after the treatment. The same measurements are done in a control condition to which the treatment is not applied. If possible, the control condition should match the experimental condition on relevant aspects.

Regarding internal validity, the major strength of this design is the ability to test for the threat of history (this threat implies that observed effects are due to an event other than the treatment which takes place between pretest and posttest). In case the conditions differ from each other considerably, the internal validity threat can only come from a unique historical event at the time of the intervention (which has different effects on the conditions due to differences between the conditions). The untreated control series also allows tests of the other threats to internal validity that operate on the single time series, such as maturation, instrumentation, statistical regression, etc. (Cook & Campbell, 1979, p. 214-216).

The effectiveness of the ProMES design is judged by its contribution to attainment of the company's objectives. In the participation phase, the evaluation of the effectiveness of the design mainly focuses on productivity changes that result from implementation of the 'prototype' and implementation of the 'final design' in the two regions that participated in the design of the system. ProMES data from the other 12 regions is used as control data. Additional evidence of a more anecdotal nature is sought regarding attainment of secondary objectives of providing technicians with useful feedback and contributing to a more accurate performance appraisal.

3.4.2. Second phase of the study: transportability of the system

The second main research question of the study is:

> ***"In this setting, can the design be successfully implemented in other –comparable–groups, without going through the participative design process?"***

In order to adequately assess the effects of transporting the system, six regions are selected as 'transportation regions' while six regions remained control regions. The transportation and control groups are matched on the average ProMES effectiveness scores in the period before the start of the transportation process. The transportation phase is illustrated in Figure 3.3 and Table 3.3.

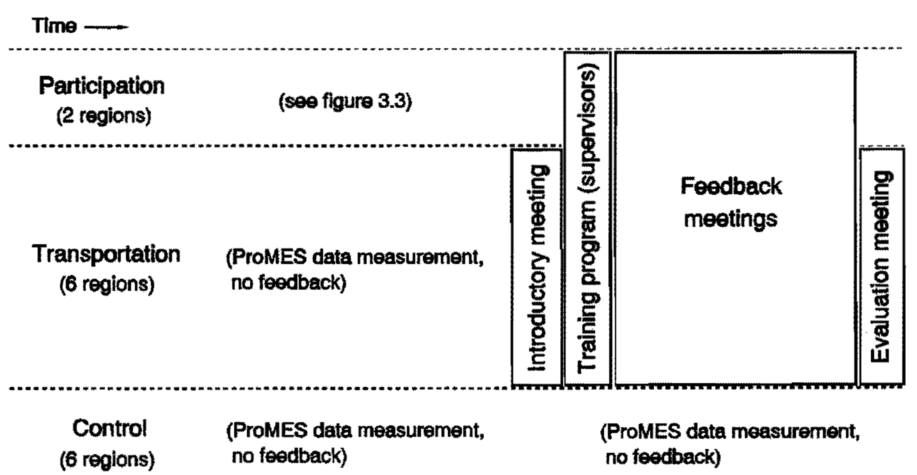


Figure 3.4. The transportation phase: a chronological overview.

Table 3.4. The transportation phase: overview of activities.

Activities	condition	Contents
Introductory meeting	transportation	A basic introduction into the ProMES system is given during a standardized introductory meeting.
Training program (supervisors)	participation & transportation	Supervisors from both conditions take part in a training program aimed at conveying to them the necessary knowledge and skills to conduct effective feedback meetings.
Bilateral feedback meetings	participation & transportation	The 'final' ProMES system is implemented in both conditions.
Evaluation meeting	transportation	The transportation process is evaluated.

To test the effects of the transportation of the system, the 'interrupted time series design with a non-equivalent control group time series design' is used once again. In this second phase of the study, there are three conditions whose effects are compared, viz. participation, transportation, and control:

participation condition $O_1\ O_2\ ...\ O_n \quad X_P \quad O_{n+1}\ O_{n+2}...O_{n+m}$

transportation condition $O_1\ O_2\ ...\ O_n \quad X_T \quad O_{n+1}\ O_{n+2}...O_{n+m}$

control condition $O_1\ O_2\ ...\ O_n \qquad O_{n+1}\ O_{n+2}...O_{n+m}$

The effectiveness of the transportation process is judged in two ways. First, the productivity change in the transportation condition is compared to that in the control condition and the participation condition. Secondly, subjective reactions to the system in the transportation condition are compared to those in the participation condition.

3.4.3. Evaluation of the designed system

At the end of the transportation phase, an evaluation of the designed system is undertaken, focusing on fulfillment of the criteria for an effective control loop:
- To what extent does the design meet the general design criteria for feedback and goals?
- To what extent have the requirements from the ProMES approach been fulfilled?
- Have effective solutions been implemented for the relevant situational constraints?

These are the main questions in an evaluation step in the design cycle in which the extent to which the design satisfies the design criteria is assessed and ideas for improving the design are generated. Apart from contributing to a more effective design, the knowledge generated in this evaluation step may serve as input for a 'pragmatical knowledge-base' for use in future ProMES projects.

This evaluation is done at the end of the transportation phase, because this enables an evaluation of the ProMES design in its final form in both the participation condition and the transportation condition.

3.5. DEFINING THE 'DEGREE OF PARTICIPATION' IN THE TWO EXPERIMENTAL CONDITIONS

In this section, we will differentiate the 'participation'-approach and the 'non-participative implementation (transportation)'-approach by placing them on a 'participation-continuum'. Furthermore, the approaches to be taken will be compared to decision-outcome criteria from the contingency models of leadership.

3.5.1. A contingency model

Degree of participation: a continuum

The degree of participation in decision-making is usually presented as a continuum which ranges from zero influence to large influence of subordinates on the decision to be made (see also Section 2.3.1). We will use one of these models–a five-point ordinal scale of decision-making strategies used by Heller (1971)–as representative of a large number of comparable contingency models. According to this model, five levels of influence of subordinates can be discerned:

1) *Own decision without explanation:* The decision is made by the leader without prior consultation with subordinate(s). The decision is not explained to the subordinates.
2) *Own decision with explanation:* The decision is made by the leader without any previous consultation with subordinate(s). The leader gives a formal post-decision explanation of the reasons for the decision.
3) *Prior consultation with subordinate(s):* The decision is made by the leader after consultation with one or more subordinates. The leader's decision may or may not reflect the subordinates' influence (it usually will).
4) *Joint decision-making with subordinate(s):* The decision results from a process of consensus formation in which one or more subordinates participate and in which some determination of a majority position is made. More often than not, the majority view is accepted, although the leader may occasionally overrule it.
5) *Delegation of decision to subordinate(s):* The leader allows subordinates to make the decision on their own. A report on the decision may or may not be requested. The leader seldom vetoes the subordinates' decision.

If one defines participation as influence-sharing between hierarchical supervisors and their subordinates (Mitchell, 1973), then strategies 1) and 2) are non-participative, and the degree of participation increases from strategy 3) to 5).

Decision-outcome criteria

Taking the contingency approach to assessing what degree of participation is appropriate, three classes of outcomes of a decision-making process can be discerned which influence the ultimate effectiveness of the decision (e.g. Siegel & Lane, 1982).

1. The *quality* or rationality of a decision.
2. The *commitment or acceptance* of subordinates to execute the decision effectively.
3. The *amount of time required* to make the decision.

The degree of participation should be based on an assessment of the quality, acceptance and time requirements for the decision. The time requirement is mainly used in case the quality and acceptance requirements produce more than one feasible strategy. The alternative to be selected is usually the one that minimizes the time expended (e.g. Vroom & Yetton, 1973).

3.5.2. Defining the participation and transportation condition

Within the context of the research setting, we will regard the design and implementation of a ProMES control-loop as a decision-making process, in which the 'leader' is the group of management representatives and the 'subordinates' are the unit members and the unit supervisor, who are to use the system. The preferred location on the 'participation-continuum' for the two 'experimental' conditions - participation and transportation - is established, based on the requirements for quality, acceptance, and time in both conditions.

The participation condition

Quality. The quality requirements for a ProMES control loop were discussed in Chapter 2, in which general design criteria, ProMES design criteria, and situational constraints were addressed. A high-quality design is a very important outcome of the decision-making process in the participation condition. An accurate measurement system will probably not be realized unless technicians and first-line supervisors are involved in the design procedure. They can contribute specific and detailed knowledge to the design process which others in the organization (higher management, staff) do not necessarily possess. In addition, the system should be consistently linked to organizational policy. The requisite knowledge for accomplishing this aim should be provided by higher management. In all, the quality requirement suggests that both 'consultation' and 'joint decision-making' would be feasible strategies.

Acceptance. Acceptance of the design, both by the unit and the management, is very important. Unit acceptance of a ProMES system which results from a non-participative decision-making process would generally be low, due to a low perceived accuracy of the system and limited understanding of the system by the unit members. This holds especially true in the Nashuatec setting in which low degree of trust between management and service units exists and the decision outcome might affect future performance appraisal procedures and distribution of rewards. The perceived accuracy of the system from the management's point of view primarily depends on the consistent linking of the ProMES system with organizational policy regarding the unit. This will determine management acceptance of the system and require some 'leader' influence on the decision in order to guarantee that this link is sufficiently present. In all, the acceptance requirement suggests that 'joint decision-making' would be the appropriate strategy.

Time. The amount of time required for designing the system is only a minor consideration since the design process involved no specific time constraints. Management gave quality and acceptance of the system a much higher priority than a short development period; it would be acceptable to management if the design process were to take several months, if only implementation an accepted, high-quality system were the result.

In case of group problems such as the development of a ProMES system, an additional consideration is whether the entire group should participate in the decision-making process or just involving one or more representatives would suffice. Although latter alternative probably reduces the amount of time required, it may negatively influence group acceptance of the decision. This will be the case if disagreement among group members over the appropriate solution is likely *and* interaction among subordinates to resolve these disagreements is prohibited (e.g. Vroom & Yetton, 1973).

Table 3.5 summarizes the above argument and shows that 'joint-decision-making' is the optimum decision-making strategy for designing a ProMES system in the Nashuatec context.

Table 3.5. Comparing the alternative design approaches to the decision-outcome criteria.

outcome	requirement	own	own + expl.	cons.	joint	deleg.
Quality	very important	no	no	yes	yes	no
Acceptance	very important	no	no	no	yes	no
Time	unimportant	yes	yes	yes	yes	yes
Acceptable strategy?		no	no	no	*yes*	no

The transportation condition

Quality. The quality of the design is not a relevant outcome of the transportation process, since sufficient quality should have been attained by the design process in the participative condition. The likelihood that the quality of a system which has been approved by two different regions and representatives from management can be substantially improved may be considered low. Thus, from a quality-point of view, all strategies are feasible, except 'delegation', which might lead to lower design quality due to lack of input from management.

Acceptance. Acceptance of the design is a very important outcome of the transportation process. Because colleagues have been heavily involved in the design of the control loop, the likelihood of acceptance of an imposed system in the transportation condition is higher than if the system were entirely designed by management or staff and then imposed. So, while 'joint decision-making' would be the optimum strategy to promote acceptance of the system, both 'consultation' and 'own decision plus explanation' *might* also be sufficient.

Time. The amount of time needed poses an important constraint in the transportation condition. Keeping in mind the requirement that one and the same system should be designed for use by all 14 regions, the benefits of 'repeating' the time-consuming design process in each new region (perhaps some quality improvement, increased acceptance) would be outweighed by the drawbacks (very long overall development time, high cost, extensive coordination between design teams). For this reason, both 'own decision' and 'own decision with explanation' would be acceptable strategies, and 'consultation' may also meet the time requirement.

Table 3.6 shows that both 'own decision with explanation' and 'consultation' may constitute acceptable strategies for 'transporting a ProMES' system in the Nashuatec context. However, there is a trade-off between the acceptance and the time criterion which reflects the dilemma faced in the second part of the project: almost certain acceptance of the system can only be obtained at the expense of an unacceptable investment of time, whereas alternatives that meet the time requirements might not result in the degree of acceptance needed.

Table 3.6. Comparing the transportation alternatives to the decision-outcome criteria.

outcome	requirement	own	own + expl.	cons.	joint	deleg.
Quality	unimportant	yes	yes	yes	yes	no
Acceptance	very important	no	yes??	yes??	yes	no
Time	important	yes	yes	?	no	no
Acceptable strategy?		no	yes?	yes??	no	no

Lastly, Table 3.7 places the decision-making strategies chosen in the participation-continuum.

Table 3.7. Location of 'participation' and 'transportation' on the continuum.

non-participative strategies		participative strategies (low → high)		
own decision (no explanation)	own decision (explanation)	consultation	joint decision-making	delegation
	transportation		*participation*	

The participation phase of the project is the topic of Chapter 4. In Chapter 5, the transportation of the system in discussed. Chapter 6 contains the evaluation of the design.

Chapter 4

Participative design of a ProMES system in two regions

4.1. INTRODUCTION

In this chapter, the elaboration of the first research question

"How can an effective ProMES control loop be designed in the Nashuatec setting with its specific characteristics?"

is presented.

The participative design and the implementation of a ProMES control loop will be described in detail, with special attention to the way various design problems were addressed (see Figure 4.1). Sections 4.2 to 4.4 describe the design process and the preliminary system (the 'synthesis' step in the design cycle). Section 4.5 deals with the (pilot) implementation of the system prototype. Its effects (the 'simulation' step in the design cycle) are discussed in Section 4.6. This is followed in Section 4.7 by a participative problem analysis, an evaluation step in which the prototype was compared to the list of requirements. This analysis resulted in a number of changes in the design and its implementation.

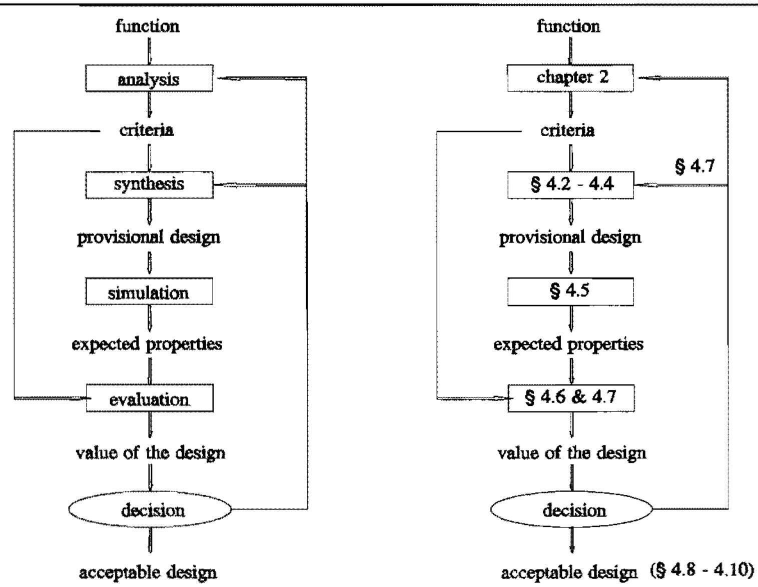


Figure 4.1. Participative design of ProMES: steps of the design cycle.

The productivity effects of the second implementation are discussed in Section 4.8. The final acceptable design is described in Section 4.9. Lastly, general conclusions are drawn in Section 4.10, based on a first re-implementation of the improved design.

4.2. PRELIMINARY ISSUES

4.2.1. Choice of design regions

In selecting the two regions that were to participate in the design of the ProMES system, four criteria were used. First, there should be a high probability of success. Therefore, the supervisors of the regions selected should have a positive attitude toward the ProMES program and its development in their regions. Secondly, existing resistance toward the program and distrust among service technicians regarding managements' intentions with ProMES should be confronted in an early stage. Keeping in mind the objective of eventually implementing the system in all regions, it would be unwise to develop the system in two very cooperative regions, only to encounter all kinds of (unexpected) problems in transferring the system to other, more critical, regions. It was therefore decided to choose two regions that differed with regard to their attitude toward management and a system such as ProMES: one (rural) region was considered cooperative, the other (urban) region was considered critical. Thirdly, the two regions selected should have some prestige with the other regions. This might facilitate the acceptance of the system in the latter regions. Finally, the average productivity of the two regions selected should be comparable to the average national productivity. In that case, the effects obtained in the implementation phase could be a fair approximation of the productivity increase that could be achieved nationally. Neither of the two regions selected should have an inferior productivity, for this would suggest a 'punitive expedition'.

Keeping these criteria in mind, two regions were suggested by the field service management:

1. 'Enschede', a rural region expected to have a cooperative attitude toward the program;
2. 'Utrecht', an urban region expected to be (very) critical toward the program.

When ProMES was first introduced to the supervisors of all regions, the supervisors of Enschede and Utrecht were among those who expressed a willingness to have their regions participate in the development of ProMES.

It was decided to develop ProMES in each of the two regions separately, in monthly meetings during paid overtime hours, with design teams consisting of all the technicians of the region (18 in Enschede, 21 in Utrecht), the region's supervisor and two facilitators.

4.2.2. Composition of the design team

Typically, the design team consists of a small group of representatives (3-5 persons) from the unit for which the ProMES system is designed, the unit's supervisor, and one or two facilitators (5-8 persons in all). The members of the design team go through the steps to develop the system, they have discussions to generate measures, they keep the other members of the unit informed, and they implement the system when it is completed (Pritchard, 1990).

For two main reasons, the decision was made to include all the region's technicians in the design team. The first reason was the highly isolated nature of a technician's job, which made communication between job incumbents during working time simply impossible. So, the process of informing all unit members of the progress made by the design team and reaching agreement on the choices made would be a very difficult one when using a design team of a few unit representatives. Achieving consensus between the design team and the rest of the unit would in that case probably necessitate a number of extra meetings. Secondly, it was felt that all technicians who expressed a desire to take part in the design of the system should be allowed to do so, especially since the system was expected to measure the productivity of individual technicians.

This decision resulted in large design teams (per region about 15 technicians, a supervisor, and two facilitators). A large group size typically involves a high complexity of interactions and communication, which make it more difficult to achieve agreement. Using group discussion as the only means of decision-making would take up too much time in large groups. For this reason, the additional use of several non-interactive decision-making techniques was necessary. These techniques are discussed in Section 4.3.

4.2.3. Introduction of the ProMES method

In both design regions, the ProMES program was introduced at a plenary meeting. First, results were reported from a questionnaire on the use of feedback and goal setting in the field service department (Algera & Van Tuijl, 1990; Kleingeld, 1990; Van der Putten, 1988), which had been completed by a large number of service technicians. Special attention was drawn to the fact that the technicians felt that the quality of feedback left a lot to be desired. Secondly, the importance of the productivity improvement program for the company was explained and discussed. Next, the ProMES method was presented. A simplified version of the ProMES system developed by the US Air force Electronic Maintenance Unit (Pritchard, 1990) was used as an example. Finally, the technicians were formally asked to participate in the development program on a voluntary basis. Technicians willing to do so would receive payment for the overtime hours they would spend on ProMES design meetings (designing the system during company hours would not be feasible, because that would cause an unacceptable backlog of calls).

In the Enschede region, the technicians adopted a wait-and-see attitude. They were willing to give the program the benefit of the doubt and start the development of the ProMES system, but they claimed the right to withdraw from the program if management would betray their trust or if the group would become doubtful of the usefulness of the approach. The reactions in region Utrecht were far more critical. Some of the technicians expressed a deep-rooted distrust toward the management, including the region supervisor. After an extensive discussion, the group was prepared to start with the program on the condition that management would give a written guarantee that the system to be developed would not be used for performance appraisal or reward purposes until all parties concerned (i.e. both management and technicians) would agree that the system was reliable and accurate enough for performance appraisal purposes. After this guarantee was given, the development process in Utrecht was started.

In summary, the following design problems, arising from the main situational characteristics, were addressed in the preliminary stage of the design process (Table 4.1):

Table 4.1. Design problems addressed in the preliminary stage

design problem	'solution'
"Individual task" A (mainly) individual performance management system requires individual acceptance. *"Lack of communication"* Infeasibility of design by small group of representatives (opportunities for communication and consensus with rest of unit prevented by geographical dispersion of unit members).	The two design team consist of all technicians willing to participate in the design process.
"Lack of trust" - General fear of being manipulated - Fear that invalid system would be used for performance appraisal	- Distrust is confronted at an early stage by having as especially distrustful region participate in the design process. - Both trial regions have the opportunity to withdraw from the project. - A written guarantee is given that ProMES will not be used for performance appraisal purposes until all parties concerned consider the system valid.

4.3. THE DESIGN PROCESS: DECISION-MAKING TECHNIQUES USED

4.3.1. The general design process: discussion until consensus

According to Pritchard (1990), the basic process used in developing a ProMES system is discussion until consensus is reached. The design team discusses the issue at hand until there is general agreement on the solution. If possible, unanimity should be obtained on the major issues. A strong majority consensus should be reached on the other issues (both through interactive discussion) (Schoonen, 1993). This interactive discussion, in which all group members should have the opportunity to participate fully, is considered essential to the success of the method. Different individuals in the unit will have different ideas on the responsibilities of the unit, which measures should be used, etc. These different ideas are typically a result of looking at the work from different perspectives, and usually all have some validity. Exposing all the members of the design team to these different views broadens the perspective of all the members and results in a better system (hence: 'constructive disagreement is good'). Pritchard therefore advises against decision-making techniques that decrease face-to-face discussion.

4.3.2. Use of additional decision-making techniques

Interactive group discussion, as advocated by Pritchard (1990), was considered an important element of the design process in the Nashuatec setting. The large design team and the limited opportunities for communication between team members, however, necessitated the use of additional decision-making techniques.

The objective was twofold: first, to keep the design team members involved during the month-long period between design meetings, and secondly, to use the time available during these meetings efficiently. Elements from three techniques were used: the Nominal Group Technique (NGT), the Delphi Technique, and scaling techniques.

Nominal Group Technique (NGT)

The Nominal Group Technique was developed as a method of group decision-making by Van de Ven and Delbecq in the late 1960s (Delbecq et al., 1975). The process of decision-making in NGT includes the following steps, which can be divided into three 'brainstorming' steps and one to three 'judgement' steps (e.g. Delbecq et al., 1975; Rohrbaugh, 1981):

1) silent generation of ideas in writing;
2) round-robin feedback from group members to record each idea in a concise phrase on a flip chart;
3) discussion of each recorded idea for clarification and evaluation;
4) (optional:) individual voting on priority ideas through rank-ordering or rating by secret ballot;
5) (optional:) brief discussion of the preliminary vote;
6) final individual voting through rank-ordering or rating by a secret ballot with the group decision being mathematically derived.

With NGT, several process losses which occur in interactive group discussion are reduced, such as the domination of the discussion by one or two influential individuals, the social pressures placed on dissenting minorities to conform to group norms, the tendency for groups to pursue a single train of thought for long periods, and the tendency to reach speedy solutions before all problem dimensions have been considered (e.g. Van de Ven & Delbecq, 1971).

The research suggests that different phases of problem-solving require different group process strategies. The Nominal Group Technique has been shown to be most effective for generating information and fact-finding concerning a problem. NGT is consistently superior to interactive group decision-making processes in generating ideas (number of ideas, number of unique ideas and quality of ideas) (e.g. Delbecq et al., 1975; Mahler, 1987; Van de Ven, 1974). However, in the 'judgement' phase, interactive groups may be more successful, because interacting group processes stimulate individuals to consider other dimensions of a problem and help synthesize and evaluate solution possibilities (Van de Ven & Delbecq, 1971). Further, interacting groups are functional in elaborating, modifying, and working toward a consensus when implementing a solution (Delbecq & Van de Ven, 1971). Also, NGT neither permits the necessary synthesis nor provides a way to deal with interrelationships among issues (Hart, Boroush, Enk & Hornick, 1985). Interaction and *constructive* conflict create the dynamics for uncovering assumptions which may lead to the synthesis of proposals or the discovery of new alternatives (Schweiger, Sandberg & Ragan, 1986). The process by which group interaction in the 'judgement' phase is structured, however, is critical in reducing the detrimental effects of interacting groups (noted earlier) which inhibit individual judgements.

While there is ample evidence for the superiority of NGT in generating ideas, evidence for two other propositions of NGT advocates, that NGT provides a greater sense of participation and (consequently) commitment to the choices made and that NGT creates consensus, is mixed. Mahler (1987) found that interactive group discussion caused a greater sense of participation than NGT. She found no significant differences with regard to acceptance and consensus. Rohrbaugh (1981) compared NGT and 'Social Judgement analysis (SJA)', an interactive group technique which provides feedback to group members about the reasons for their disagreements. The SJA-technique was shown to be superior to NGT in creating consensus because NGT limits the examination of the reasons for conflict.

Use of the Nominal Group Technique in developing ProMES

While NGT is shown to be superior in generating more high quality ideas than interactive group processes, the boundaries imposed by NGT on interaction and disagreement may limit group members' sense of participation and the synthesis of ideas into a final choice. This conclusion led to the decision to use NGT in the 'brainstorming' phase of the ProMES design steps for products and indicators (in these large design teams, using NGT in the generating phase turned out to be very effective: in one instance, as many as 36 potential indicators were generated). In the 'judgement' phase interactive discussion was used, guided by the criteria for products and indicators (which represents a higher degree of interaction than the 'standard' application of the NGT). To facilitate these discussions, the design team was divided into two to four smaller groups. In plenary sessions, the findings of the subgroups were presented by representatives from each group and discussed until consensus was reached. These discussions were structured by using the criteria for products and indicators as guidelines.

Delphi Technique

The Delphi Technique uses panels of experts or selected participants who typically never meet face to face, but who report their individual preferences among proposals. Preferences from one round of questioning are aggregated and fed back to the members for re-ranking. This continues for several rounds until convergence of judgement reaches some point of diminishing returns. The interaction of the participants as a group is limited to responding to the aggregated rankings from the prior rounds of responses. Direct persuasion or negotiation is not possible (Linstone & Turoff, 1975). The outcome feedback permits a carefully restricted exchange of information while reducing the process losses which might occur during traditional group interaction. Few controlled experiments have assessed the effectiveness of the Delphi Technique (Gustafson, Shukla, Delbecq & Walster, 1973; Rohrbaugh, 1979). They typically report unconvincing findings with regard to accuracy of judgement and reduction of disagreement.

Use of the Delphi Technique in developing ProMES

An element of the Delphi Technique was used to make efficient use of the time period between design meetings, since the geographical dispersion of the technicians prevented informal discussions about design issues between meetings. Questionnaires were sent to all design team members, asking them to rate provisional products and indicators on the extent to which they met the criteria. At the beginning of the next design meeting results were presented by the facilitators and discussed interactively.

In developing the indicators, the NGT brainstorming-sessions resulted in a large number of potential indicators. These suggestions were collected by the facilitators. The next design meeting, a checklist containing the indicators (categorized per product) and the criteria for acceptable indicators was submitted to the individual design team members for rating. The combined results of the individual ratings were a starting point for discussions about which potential indicators to use (possibly in an adapted form or in combination with another suggested indicator).

In developing the contingencies, a questionnaire was distributed among design team members to determine individual preferences for alternative shapes of the contingencies. Their individual preferences were used as a starting point for group discussion.

As in the application of NGT, the structured part of the decision-making process was limited to the generating phase. Taking individual ratings of the suggestion as a starting point, interactive group discussion was used to achieve consensus on the final choices.

Scaling techniques

An important element of the ProMES design process is the determination of the relative importance of the indicators. Pritchard (1990) suggests a two-step procedure. First, maximum indicator values are ranked in terms of their contribution to the overall effectiveness of the unit. Secondly, the group rates the maxima relative to the +100 effectiveness score of the most important maximum. Analogously, the minimum effectiveness scores are determined.

When the number of indicators is large, these comparison processes become rather complex, especially when they are carried out in a large group. To reduce this complexity, two individual scaling techniques were used in the Nashuatec design process. First, through the technique of paired comparisons (Edwards, 1957), the simultaneous comparison of 10 to 14 indicators was divided into a number of simple comparisons of two indicators at a time. This reduced the complexity of the comparison process. The results of these comparisons, carried out individually by all design team members, were used as starting point for group discussion. In addition, an individual ranking technique was used. The application of these techniques will be discussed in more detail in Section 4.4.3 (Establishing the contingencies).

Table 4.2 summarizes the techniques used in each of the ProMES design steps at Nashuatec. Table 4.3 shows which design problems were addressed by each of the techniques.

Table 4.2. Decision-making techniques used in the design process at Nashuatec.

Design team: all technicians willing to participate in the design process	
design step	decision-making techniques used
products	Nominal Group Technique (generating stage) Delphi Technique (questionnaires) discussion in small subgroups & plenary group discussion until consensus
indicators	Nominal Group Technique (generating stage) Delphi Technique (questionnaires & checklists) discussion in small subgroups & plenary group discussion until consensus
contingencies	Delphi Technique (questionnaire) scaling methods (paired comparisons & ranking) plenary group discussion until consensus
feedback report	plenary group discussion until consensus

Table 4.3. Design problems addressed

Decision-making technique	Design problem 'solved'
Nominal Group Technique (generating stage)	- large design team: more efficient use of time more effective generation of ideas
Delphi Technique (questionnaires)	- no possibilities for informal communication between meetings: effective use of time between meetings
Scaling techniques (ranking/paired comparisons)	- large design team - large number of indicators

4.4. SYSTEM DESIGN

4.4.1. Developing products and indicators

Products

Products were defined by the facilitator as: "the essential contribution of a service technician to the company". This definition was paraphrased to aid the technicians in generating ideas: "What are my responsibilities as a technician?"; "Why did Nashuatec hire me?", and "What is my job as a service technician?". The following criteria (comparable to Pritchard, 1990) were used to determine which potential products would be suitable:

- The product is the responsibility of service technicians, not of others in the company.
- The product is relevant for the company, i.e. the company benefits if service technicians do exactly what the product says.
- The list of products is complete, i.e. it covers all important aspects of a technician's job.

The products were generated by the design teams according to the procedures described in Section 4.3. Initially, seven products were generated in Enschede and nine in Utrecht. During the discussions that followed, several of these potential products were rejected. For example, the proposed product 'Selling paper' (technicians were supposed to turn the client's attention to the fact that the company was selling paper for use in copiers too) was rejected, because technicians considered this an activity which had nothing to do with their core responsibilities (repairing and maintaining photocopiers) on which they should spend their time and effort. Another product, 'Keeping technical knowledge up-to-date', was removed from the final set of products, because other departments were also responsible for this, and because the effects of up-to-date technical knowledge would become apparent in other products, for which better indicators could be developed. See Table 4.4 for the final set of products.

Indicators

Indicators were defined as: "Concrete measures, which show how well a technician generates a product". Questions asked to help the design team generate ideas were: "How would you show your supervisor how well you perform on this product (or: that you were doing a good job on this product)?"; " How could you measure how well you do your job on this product". Six criteria were used to assess the appropriateness of the possible indicators.

1. The set of indicators should be complete, i.e. every product should have at least one indicator, and all different aspects of a product should be covered by indicators.

2. The indicators should be valid, i.e. what is measured should be an accurate representation of product accomplishment. This includes long-term effects which should be neutral or positive.
3. The indicators should be controllable.
4. Data on the indicator should be cost-effective to collect.
5. The indicators should be understandable and meaningful to the technicians.
6. There should not be a large overlap between the indicators (they should not measure the same things).

Using the Nominal Group Technique, many potential indicators were generated (20 in Enschede, 36 in Utrecht). Using a checklist, these suggestions were first rated against the criteria for validity and controllability. The other criteria were then applied to the remaining indicators (15 in Enschede and 24 in Utrecht). The final ten indicators conformed to all criteria.

There was a lot of discussion regarding the third criterion: controllability of the indicators. For most proposed indicators, numerous factors, apart from the technician, influenced the results that could be obtained. These factors include the customer (how is the copier handled, how many copies are made per month, where is the copier located), colleagues (did the previous technician perform incomplete repair or maintenance), the planning department (which machine is assigned), and the warehouse (are the right spare parts delivered in time). It was clear that a hundred percent controllability could not be achieved; therefore, the minimum controllability requirement was defined as: a technician should be the most important factor of influence and should therefore control at least 50 percent of the outcome variability.

The final set of indicators, agreed upon during the final review and approval meeting with the management, is listed in Table 4.4.

Review and approval meeting (products and indicators)

The discussions with the management of the field service department were intense and valuable. This was partially because for many technicians this was the first real contact with the managers. Management expressed their appreciation of the work done by the design teams. Several important issues were settled in direct interaction. For example, management insisted on technicians wearing official company work clothes. However, these work clothes were not delivered on time, so that technicians often had to do their job dressed unofficially. Similar contradictory requirements had to do with the delivery of spare parts. Management asked for low percentages 'Return calls due to incomplete car stock', but spare parts were sometimes out of stock, or the warehouse made mistakes in supplying the technicians. It was not possible to resolve all these issues, but they were brought to the attention of management and taken seriously by them. The controllability criterion was discussed extensively here. Responsibility was only accepted by the technicians in cases where solutions could be given or promised for the near future. One general solution offered was to use running averages over longer periods as measures in addition to scores per month.

Table 4.4. Products and indicators.

Product and Indicator(s)	Definition of the indicator	Type dependent (see 4.4.2)	Level of measurement (see 4.4.2)
Product 1. Quality: to repair and maintain photocopiers as effectively as possible.			
Indicator 1.1: Mean Copies Between Calls	- The average number of copies made between a technician's repair visits and the first malfunctions after those visits.	yes	individual
Indicator 1.2: Percentage repeat calls	- The percentage of repair visits made by a technician within five working days of the original repair.	yes	individual
Indicator 1.3: Compliance with preventive maintenance procedures	- The percentage of preventive maintenance prescriptions correctly followed.	no	group
Product 2. Cost: to repair and maintain photocopiers as efficiently as possible.			
Indicator 2.1: Parts cost per call	- The average amount of money spent on replacement parts.	yes	individual
Indicator 2.2: Labor time per call	- The average amount of time used for repairing and maintaining photo copiers.	yes	individual
Indicator 2.3: Percentage return calls due to car stock	- The percentage of visits caused by a lack of spare parts in the technician's car stock.	yes	individual
Product 3. Administration: to keep records of repair and maintenance activities as accurately as possible.			
Indicator 3.1: Accuracy of History Card	- The percentage of required information filled in correctly on the History Card.	no	group
Indicator 3.2: Completeness of claims	- Percentage of potential claims (parts replaced within their warranty period) submitted.	no	individual
Product 4. Attendance: to spend the available time on work related activities.			
Indicator 4.1: Percentage of capacity used	- The percentage of labor contract hours actually spent on the job.	no	individual
Product 5. Ambassadorship: to behave as correctly as possible on the job.			
Indicator 5.1: Correctness of behavior	- The percentage of important social behaviors demonstrated on the job.	no	group

By taking averages, the influence of chance factors resulting in accidentally high or low scores is minimized, thereby giving a more accurate picture of a stable 'true' score. A specific solution concerned the indicator 'Parts cost per call'. Technicians were willing to take responsibility for this indicator, provided that the replacement intervals in the maintenance procedure would be abolished or lengthened, thereby increasing the technicians' latitude in replacing spare parts.

It should be mentioned that the discussion with the management also resulted in the addition of an indicator. Indicator 1.3, 'Compliance with preventive maintenance procedures', was not in the list generated by the design teams. In discussing the products and indicators proposed to management, management felt that improving performance on the proposed indicators could be beneficial to the company in the short run, but disadvantageous in the long term. The reason was that preventive maintenance could easily be neglected in favor of labor time and parts cost per call. The net short term effect would be high scores on these indicators and acceptable scores on 'Mean Copies Between Calls' and on 'Percentage repeat calls', both short term quality indicators. A probable long term effect would be a decrease in the general condition of the photocopiers, eventually resulting in a high rate of machine malfunctions. To prevent this, a new indicator needed to be developed which measured preventive maintenance quality, consisting of a checklist covering the main elements of the preventive maintenance procedures for each type of photo copier at several points in the machine's history.

Of all ten indicators, only the four indicators 'Compliance with preventive maintenance procedures', 'Accuracy of History Card', 'Correctness of behavior', and 'Percentage of capacity used' were newly developed indicators by the design team. The company already gathered information on the other indicators, but there were doubts on the validity and meaningfulness of that information. The main reason for the latter point was that data concerning different types of photocopiers, i.e. high volume, middle volume and low volume machines, were not comparable. Even though this was a well-known problem, the company simply added the scores across these types of copiers, resulting in feedback that was not meaningful. As will be shown below, the contingency technique offered by ProMES can solve this problem: seemingly incomparable scores on identical indicators can be added in the same way as scores on different indicators.

4.4.2. Operationalizing the indicators

The operationalization of the indicators turned out to be a very time-consuming and complex activity. In between development meetings, the facilitators worked on the indicators with the help of product support specialists, supervisors, and a system analyst. The results were reviewed by the design team during design meetings.

Level of measurement

An objective of this ProMES system was to measure the productivity of individual service technicians. If possible, all indicators should therefore be measured at the level of individual technicians. Measurement of group performance–a source for a second type of feedback aimed at preventing competitive behavior which would be dysfunctional in case of interdependence between technicians–should be obtained by aggregating the individual results. For three indicators, however, it turned out to be too costly and time-consuming to come up with measures which could provide reliable short-term feedback. For that reason, it was decided to measure individual performance on these indicators by sampling and to use the mean sampled scores as group scores. See Table 4.4 (column 'level of measurement').

Type dependency

It was found that five indicators were 'type-dependent'. That is, technical characteristics of each type of copier determined to a large extent the possible performance levels on these indicators. For example, historical data showed that on model X a MCBC of 15,000 copies is an excellent result, whereas on model Y it is just average, and on model Z it is very bad. This has far-reaching consequences for the development of contingencies, because, for these five indicators, a separate set of contingencies had to be developed for each of the 26 types of copiers included in the system.

Some indicators turned out to be similar to measures already in use in the service department. However, most existing measures were not suited for measuring productivity on an individual level, did not take into account differences between types of copiers, or included elements the technicians could not control. The existence of these measures meant that much of the raw data needed for the ProMES indicators was already being reported by the service technicians on a daily basis (e.g. how long did the visit take, which spare parts were used, how many copies were on the copy counter). Hence, the operationalization focused on redefining the indicators and combining the data. All definitions refer to a measurement period of one month, which was customary in the service department. Providing feedback more frequently would not be feasible.

There were ten indicators in all, three of which are measured type-independently at the group level (Compliance with maintenance procedure, Accuracy of History Card, and Correctness of behavior). The other seven indicators were measured individually. Five of these were type-dependent (MCBC, Percent repeat calls, Labor time per call, Parts cost per call, and Percentage return calls due to car stock). The remaining two indicators, Completeness of claims and Percentage of capacity used, were type-independent. See Table 4.4 for these distinctions.

Difficulties in operationalizing the indicators

For a few indicators, the operationalization process was rather straight-forward. Most indicators, however, posed serious difficulties, three of which will be discussed.

The first problem is related to the existence of *return calls*, visit(s) following the initial visit that are necessary to complete repairs and maintenance. These occur because the technician lacks spare parts or the time or knowledge to complete the initial call. In about 10 percent of the cases, the return call is handled by another technician, which complicates the allocation of labor time and spare parts usage to the individual technicians involved. Keeping in mind the objective of measuring individual productivity, it was decided to exclude calls involving more than one technician from the set of calls that is used to calculate parts cost per call and labor time per call. In that way the results of individual technicians are not influenced by their colleagues.

The second complicating factor is a *time-lag* that sometimes exists before indicator values can be calculated. An example is Mean Copies Between Calls (MCBC). In order to get this measure, the counters are compared for two consecutive calls on a machine, the difference being the number of copies made between the two calls. This is attributed to the technician involved in the first of the two calls. Averaged across calls, the MCBC is calculated per type of copier. So, it is not possible to calculate the number of copies between calls until the next call has occurred. This makes it difficult to calculate the MCBC for a technician at the end of a month, because it is very likely that not all of his repairs in that month have resulted in a new call. In fact, before the end of the month, only the less successful repairs will have resulted in new calls, the MCBC of which would not be an accurate measure of the technician's performance. For this reason, the design team decided to attribute MCBC results to the month in which they can be calculated. This is the only way to arrive at complete information, although part of the MCBC results will have originated from months prior to the feedback period.

Thirdly, it was found that there were no existing measures available for the indicators 'Compliance with preventive maintenance procedures', 'Accuracy of History Card' and 'Correctness of behavior'. Although some technicians were skeptical about introducing subjectivity into the system, the consensus eventually was that the region supervisor would be the person most capable of measuring to what extent the technicians met the demands posed by these indicators. He could do that on two occasions: during a visit, by watching the technician at work, or after a visit by looking at the results of the work done by a technician the previous day. Yet, on practical grounds it would be impossible to get an accurate monthly measure of all 20 technicians' individual performance. This is unfortunate because two of these indicators reflect the interdependence between the technicians. For example, a technician could obtain a high effectiveness score on labor time per call by carrying out only part of the maintenance procedure. This must be avoided, because it negatively influences the long-term quality of the machine and causes extra work for his colleague who visits the machine next. It was decided to measure the above indicators by sampling, resulting in group level indicator values that are allocated to the region as a whole *and* to all individual technicians. This again emphasizes the fact that the interdependencies between technicians are mainly reflected in the indicators 'Compliance with preventive maintenance procedure' and 'Accuracy of History Card' and that there is a joint responsibility to perform well on these indicators. Since differences in the extent to which individual technicians interpret and adhere to maintenance and History Card instructions would automatically cause differences in labor time spent and parts replaced, strict adherence to these instructions can also be considered a prerequisite for the validity of the type-dependent indicators.

Most of the work on the group indicators was done by the facilitators. Critical Incident interviews were conducted with supervisors, field service management and product support specialists. This resulted in four checklists: two checklists for the 'Preventive maintenance' indicator to be used during a visit and after a visit respectively, one checklist each for the 'History Card' and 'Correctness of behavior' indicators to be used after the technician's visit. Also, a procedure for applying these checklists was developed. See Appendix A for the checklists used (translated from Dutch).

Evaluation of the operationalization phase

The above description of the operationalization of the indicators shows that the amount of participation within a design process may vary considerably, depending on the specific phase. Whereas the technicians were heavily involved in establishing the products and indicators, their involvement in the operationalization process was limited. Latter process was characterized by a general lack of involvement of the technicians in the design team. Most major decisions were made during the development meetings, but the facilitators (and in some cases, the supervisors and the information systems expert) rather than the technicians did most of the thinking and deciding. Although this was the only procedure possible, there was a risk of causing limited insight into the system, low perceived validity and, as a consequence, low commitment on the part of technicians. There indeed was a temporary lack of insight into the system: during the last development meeting and the first pilot feedback meetings, the facilitators were called upon to explain how certain indicators were operationalized. Most doubts about the validity of the indicators could be taken away when the facilitators explained why certain choices were made. An important limitation in this respect was the fact that most indicators were not 100 percent controllable by individual technicians. The indicators were operationalized in such a way as to exclude as many uncontrollable factors from the indicators as possible, without diminishing their relevance. The technicians agreed with most choices that were made. Some minor changes were suggested and incorporated into the system. All in all, everyone was satisfied with the eventual operationalizations.

Although it appears to be an essential step in putting together the measurement system, not much attention is given in the ProMES literature (e.g. Pritchard, 1990) to the operationalization of the indicators as to other phases in the development of ProMES. Nevertheless, the effort involved in developing measures that remain accurate under all possible circumstances should not be underestimated[1]. This will become

---

[1] Some examples of questions that were relevant for the specific definition of the indicators are: Suppose several consecutive repeat calls on the same machine are 'caused' and 'solved' by different technicians, which repeat calls should be attributed to which technician? When a technician installs a machine, this is not considered a 'call' relevant for ProMES; if this installation is followed by two return calls and then by a repeat call caused by another technician, will latter technician get a repeat call or not? Should an MCBC-result be attributed to this technician? How should these distinctions be incorporated in the program which is to generate the feedback reports?

clear in our discussion of the pilot feedback phase, during which it became apparent that–however satisfied one had been with the indicators on paper–gathering data in accordance with the operationalizations will reveal additional problems.

4.4.3. Establishing the contingencies

The design of this stage differed in some respects from the usual approach (Pritchard, 1990). First, the first two stages in the development of the contingencies ('identifying maxima, minima, and zero points' and 'establishing maximum and minimum effectiveness scores') were switched. Secondly, scaling techniques were used extensively in the initial stages of establishing maximum and minimum effectiveness scores. Thirdly, historical data were used to establish minima, maxima, and zero points for the type-dependent indicators. These differences will be discussed.

Establishing maximum and minimum effectiveness scores

Two methods were used get a approximations of the relative importance of the indicators. First, the technique of *paired comparisons* (Edwards, 1957) was used. In applying this technique, the individual technicians were asked the following question for each pair of indicators: "Suppose you perform at the expected level (not good/not bad) on all indicators; on which of these two indicators would you want to perform maximally to maximize the value of your performance for the organization?" The results of these paired comparisons were converted into a ratio scale for maximum effectiveness scores by giving the maximum with the highest importance an effectiveness score of +100 and deciding on the effectiveness score of the least important maximum (which is probably relatively close to but above zero). The latter effectiveness score was determined by asking how effective a maximum value on this indicator would be, compared to the effectiveness of +100 for the most important maximum. The effectiveness scores in between are determined by the results of the paired comparisons. The minimum effectiveness scores were determined in an analogous way, the effectiveness score of the minimum with the highest importance being determined by the design team.

Secondly, a *ranking* technique was used in both regions, in addition to the paired comparisons[2]. Each individual technician was asked the following question: "Suppose you perform at the expected level on all indicators. What is the first indicator on which you would want to perform excellently to maximize the value of your performance for the organization ? What is the second one, third one, etc.? " An analogous question

[2] The paired comparison technique is the most powerful of the two techniques, since it enables the indicators to be expressed in an interval scale, whereas the ranking technique produces an ordinal scale. A relative drawback of the paired comparison technique is its laboriousness: it involves relatively complex calculations which cannot be done during a design meeting. For this reason, the results from the simpler ranking procedure were used until results from the paired comparisons were available the next meeting.

was asked for the minimum effectiveness scores. By averaging the rankings across technicians, a ranking of maximum and minimum effectiveness scores was calculated.

The initial effectiveness scores determined by paired comparisons and ranking were the starting point for *group discussion* on the exact positioning of minimum and maximum effectiveness scores. Two *additional criteria* were used in this discussion:

1: The relative importance of the products should be correctly reflected by the indicators.

This criterion was used because not all indicators are entirely independent. For example, there is a dependence between Mean Copies Between Calls and Percent repeat calls: if a technician succeeds in lowering his percent repeat calls then his MCBC will probably be higher. Consequently, the product being measured (in this case: Quality) might be overweighed in the measurement system. The importance of each indicator was determined by taking the range from minimum to maximum effectiveness score. This is a better way of establishing the relative importance of indicators than comparing the maximum scores as suggested by Pritchard (1990), provided the minimum effectiveness scores too are determined in a valid way[3]. In addition, problems with determining the importance of indicators having a maximum effectiveness score of zero or with indicators having asymmetrical effectiveness scores (e.g. a minimum of −65 and a maximum of +40) are avoided using this approach.

2: Maximum effectiveness on Quality should compensate for minimum effectiveness on Cost and vice versa.

The design team felt that there was a trade-off between the two products Quality and Cost. This trade-off should be reflected in the maximum and minimum effectiveness scores of the indicators belonging to those products. Thus, a technician with maximum (minimum) effectiveness on quality and minimum (maximum) effectiveness on cost should have a total effectiveness score of about zero, indicating his performance is about average. In others words, quality and costs are considered equally important aspects of a technician's performance.

In a later stage of the project, these tentative criteria were replaced by a quantitative analysis of cost reductions resulting from improvements on the indicators and a qualitative estimation of changes in customer satisfaction. For a short description of this procedure, see Section 4.7.6.

Identifying maxima, minima, and zero points

During the development of the indicators it became clear that the design team would not be able to identify maxima, minima, and zero points for most of the *type-*

[3] In the US Air Force project (Pritchard et al., 1988, 1989; discussed in Section 1.4), unit personnel had difficulty in scaling the lowest point on the indicators. This minimum possible level was difficult for them to conceptualize, and unit personnel expressed low confidence in the values set eventually. In the feedback phase of the project, the unit rarely performed near the minimum on an indicator. In the Nashuatec project, historical data on the indicator scores was available for all individual technicians; this enabled better estimation of minimum scores.

dependent indicators. Hence, it was decided to generate historical data to help identify these points. The following decisions were made in generating the data:

* The data should be collected per type of copier.
* A minimum monthly number of five calls per technician per type of copier should be used to avoid atypical results. For example, a technician performing a single call on type X might have either 0 percent or 100 percent repeat calls. Neither indicator value would be very informative.
* The data should cover a one year period to make sure any seasonal trends are incorporated into the contingencies.
* Types of copiers for which insufficient data was available to give an accurate estimate of minimum, zero point and maximum would be excluded from the system.

For all types of copiers a frequency distribution of indicator values was generated for all five type-dependent indicators.

In a discussion between management and the design teams, management stated that, in their opinion, the service technicians had performed according to expectations during the period concerned. Therefore, they suggested that the average national indicator values on all types of copiers be considered the neutral level of performance (neither good, nor bad). The design teams agreed with this point of view. The facilitators suggested that the range of results in the frequency distribution might be used to determine minimum and maximum indicator values. Management and the design team agreed, indicating that this would be the only practical way to determine these values. Figure 4.2 shows how historical data was used in establishing the contingencies.

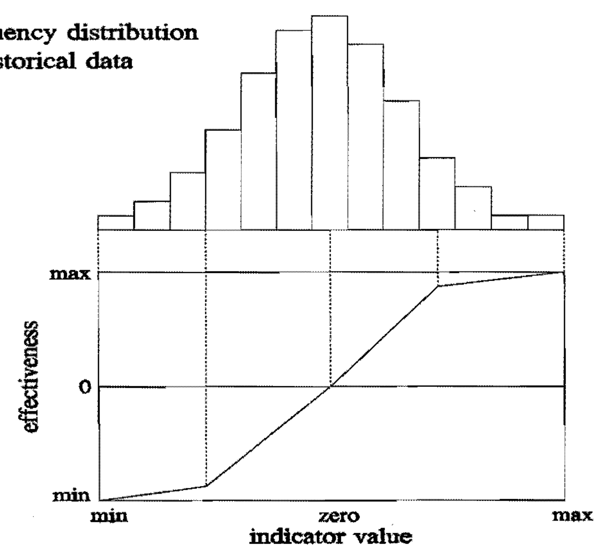


Figure 4.2. Establishing maxima, zero points, and maxima.

It has to be noted that it would probably be incorrect to base contingencies on historical data without having management and design team ask themselves whether the resulting maxima, zero points and minima, which are of a descriptive nature, can be used normatively. In the worst case this might result in contingencies formalizing ineffective and/or inefficient policies. In this organization, the service department used procedures and working methods that were quite sophisticated when compared to many companies in the industry, so there appeared to be no risk of reinforcing ineffective policy.

For 26 types of copiers, minimum, maximum and zero points were established for the five type-dependent indicators, using historical data. For the remaining 50 types (then constituting about five percent of all calls on photocopiers), no accurate values could be established due to the very small amount of historical data. These types were excluded from the system. New types of copiers were to be added as enough historical data would be available for establishing accurate contingencies.

The maxima, minima, and zero points of the five *type-independent* indicators were determined through discussion. To give an example, the neutral indicator value of 'History Card' was set at 95 percent (indicating that 95 percent of the weighted items on the History Card should be filled in correctly by the technicians examined to obtain a group effectiveness score of zero). This means that only small mistakes or omissions would be found acceptable. The minimum and maximum value were set at 80 and 100 percent, respectively.

Possible consequences of switching stages

The first two stages in the formal (Pritchard, 1990) procedure of developing the contingencies ('Identifying maxima, minima, and zero points' and 'Establishing maximum and minimum effectiveness values') were switched to bridge the time needed to collect the historical data to be used to establish minima, zero points and maxima for the type-dependent indicators. Two possible pitfalls were identified.

First, the sequence followed here does not take into account indicators which have a maximum effectiveness score of zero (i.e. equal to the expected level). In this case, a few indicators might have had such contingencies, e.g. the maintenance indicator. Yet, when the technicians' attention was called to this possibility, they held the view that they should have the opportunity to 'score some points' on each indicator; in other words, every maximum should have a positive effectiveness score. Secondly, this procedure neglects possible differences between types of copiers concerning the relative importance of indicators, requiring different strategies. For example, type X may be a relatively labor-intensive type, where labor time is a very important cost indicator, whereas type Y may be a relatively parts-intensive type where parts cost per call is a very important cost-indicator. These differences would not be reflected in the system, because the maximum and minimum effectiveness scores of the indicators are the same for all types of copiers. Although these differences between types exist, it was decided by facilitators and management not to take them into account, because it was felt that the resulting complexities could not be handled.

Drawing the complete contingencies

The objective of this stage is completing the contingencies by filling in the remainder of the points. The total number of contingencies was 135, consisting of 130 type-dependent contingencies (5 indicators for 26 types of copier) and 5 type-independent contingencies.

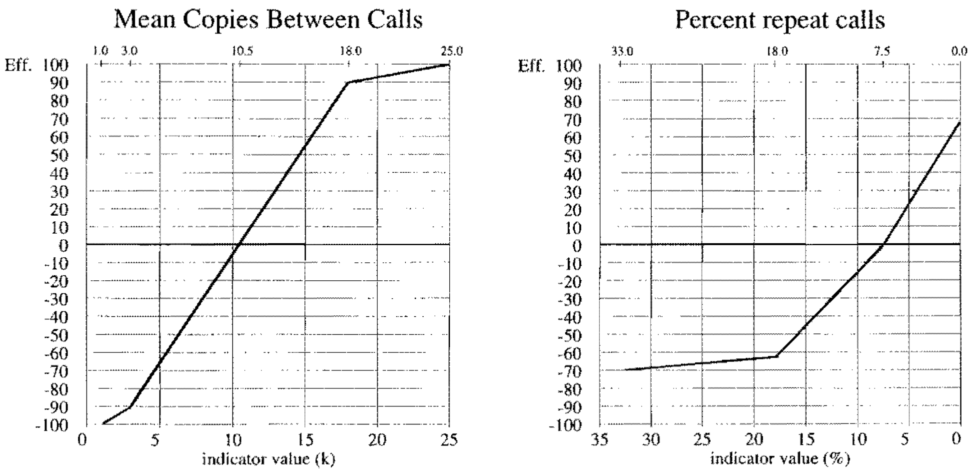
In determining the precise form of the contingencies for the *type-dependent indicators*, a well-known type was used as an example. Using transparencies, the facilitators started by simply drawing straight lines between minimum and zero point and between zero point and maximum. This caused an inflection point at the zero point for all five indicators, which according to the design team would not correctly reflect the change in effectiveness going from an indicator value just below to one just above the zero point. This change in effectiveness should, in their opinion, be about linear. Keeping in mind that most technicians generally have indicator values around the zero point, it was decided to draw contingencies which were relatively steep around the zero point and relatively flat near the minimum and maximum. The exact location of the resulting inflection points was determined using the frequency distribution of historical indicator values. The design team decided that 10 percent of the observations should be above the upper inflection point and that 10 percent of the observations should be below the lower inflection point.

After the design team had developed the complete set of contingencies for this example, the facilitators were authorized to develop the contingencies for the remaining 25 types of copier, applying the rules mentioned above. Figure 4.3 shows two type-dependent contingencies (Mean Copies Between Calls and percent repeat calls) for two types of copiers (Type 4100 and Type 7150). Note that the shape of the contingencies is roughly the same for the two types. However, the type of copier determines to a large extent which indicator values correspond to minimum, zero and maximum effectiveness.

The *type-independent contingencies* were determined through discussion. There were heated discussions regarding the attendance indicator ('Percent of capacity used'). Whereas most design team members agreed that 100% capacity used should be the neutral value (i.e. working exactly the number of hours specified in the labor contract), there was disagreement about the positions of the minimum and maximum. Some technicians favored a maximum far above zero (e.g. 110 percent), so they would be able to gain effectiveness points just by working longer. Others argued that might lead to exploitation of technicians and preferred a maximum value close to the zero point (e.g. 101 percent). A similar discussion centered around the minimum.

During the *final review and approval meeting*, management decided more or less unilaterally that an indicator value of 102.5 percent would correspond to a maximum effectiveness score of +30, and that an indicator value of 97.5 percent would correspond to a minimum effectiveness score of –80. Management argued that they expected technicians to work the amount of hours specified in the labor contract (hence, the –80), whereas they would not want technicians to overwork themselves (hence, the +30).

Type 4100: Quality



Type 7150: Quality

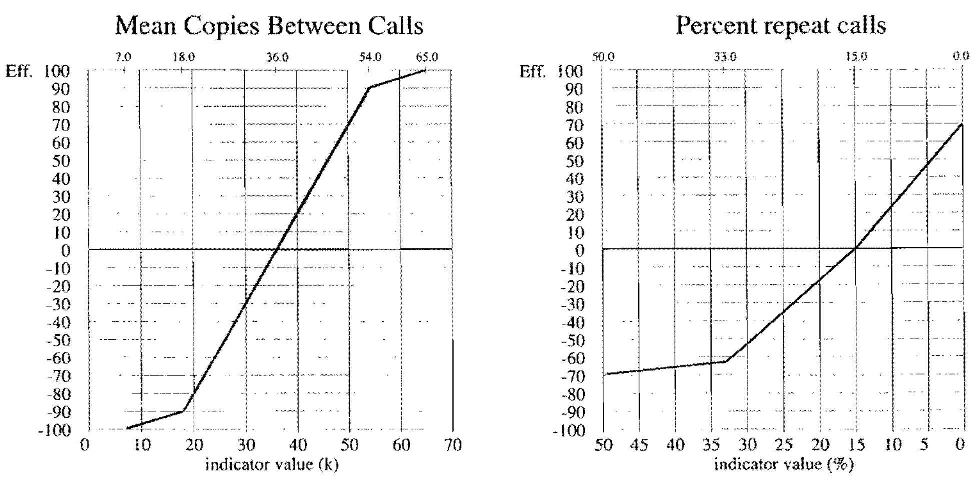


Figure 4.3. Contingencies ('Quality' indicators for two different types of copiers).

This dismayed a lot of technicians who felt that this asymmetrical valuation of indicator scores was unfair to them. This was one of the few occasions that the design team and management did not reach consensus on a contingency.

4.4.4. Designing the feedback reports and feedback meetings

Background

The ProMES system for the Nashuatec technicians is aimed at employees who work individually, but are dependent on their colleagues. As could be seen in Section 2.2.5, in designing feedback systems, interdependency between group members has to be taken into account. According to Matsui et al. (1987), in a case of high interdependence between group members, effectiveness of task feedback is maximized when it provides information on individual *and* group performance. Only providing group feedback will not cause those individuals who are below target to improve their performance, as long as the group is on target. Also, possibilities for social loafing are created. Just providing individual feedback will not cause individuals who are on target to improve their performance when group performance is below target. Mitchell and Silver (1990) argue that setting individual goals in an interdependent task results in lower performance than setting group goals, group goals plus individual goals, or no goals at all because of dysfunctional effects of competitive feelings, strategies and behavior resulting from individual goals.

Keeping in mind the interdependence between technicians–one of the main design problems (described in Section 3.2.4)–it is clear that there should not only be an individual feedback report, but also a group feedback report. From a practical point of view, a group feedback report is indispensable as a common frame of reference and a means to stimulate mutual help to improve the effectiveness of individual technicians and thereby the effectiveness of a group as a whole. In consequence, it would be also preferable to measure all indicators at the level of individual technicians. As shown in Section 4.4.2, this was possible for seven of the ten indicators.

Design process

As a basis for discussion in the design team, the facilitators prepared a first draft of an individual and a group feedback report, based on their and Pritchard's (1990) ideas on what would be useful feedback and how it might be presented. Three questions were posed by the facilitators, dealing with the feedback report, feedback meetings, and the extent to which the reports should be private or public.

1. *"What information on your own performance and on the performance of others would you like to have ?"*

After some discussion, there was agreement that every technician should have detailed information on his own performance. In addition, every technician should have information on group performance.

2. "How would you like to use the reports during the feedback meetings ?"
Initially, there was difference of opinion in the design team on how to use the feedback reports. Some technicians argued that every technician should judge for himself how to use the feedback reports. Only if the individual effectiveness score compares unfavorably to the region effectiveness score, should a technician try to improve. Other technicians argued that communication and exchange of information between technicians would be essential to improve the effectiveness of individual technicians and thereby the region effectiveness, especially for the type-dependent indicators. The knowledge and experience of technicians who perform well on these indicators should be shared with technicians who experience problems on these indicators. After ample discussion the former technicians were convinced by the latter. In this discussion the facilitators stressed that a feedback meeting should be a fact-finding exercise, not a search for excuses or a method for attaching blame. This might have put some technicians at ease who still appeared to have some fears concerning objectives and use of the feedback meetings. Finally, the design team reached agreement on the a general procedure during the feedback meetings (see Table 4.5).

Table 4.5. The suggested procedure during feedback meetings.

Type of indicator	Procedure
Group indicators: - Compliance with preventive maintenance - Accuracy of history card - Completeness of claims *Type-independent individual indicators* - Percentage of capacity used - Correctness of behavior	Discussion in full session, conducted by the supervisor. The supervisor should, among other things, present his findings pertaining to the region indicator checklists.
Type-dependent indicators: - Mean Copies Between Calls - Percent repeat calls - Parts cost per call - Labor time per call - Percent return calls car stock	Discussion in small 'type' groups (about five technicians). The groups should be discussing one or two types of copiers on which all members receive feedback. Information per type of copier, especially information on individual differences, would be necessary in order to have fruitful discussions on causes of performance differences between technicians.
Plenary announcement of the decisions and strategies arrived at in the 'type' groups and written confirmation of decisions and points for attention in the next month(s).	

3. "Who should get (parts of) the feedback reports ?"
There was agreement in the design team that the region supervisor should have access to all individual information. The general opinion was that the supervisor is responsible for the functioning of the technicians in his region. He should, therefore, be able to approach technicians individually in all matters regarding the functioning of individual

technicians. The service management should receive region feedback reports, because they are responsible for the effectiveness of the regions as a whole.

A revised draft of the feedback reports, approved by the design team, was presented to management. Management agreed with most of the design team's suggestions on how to use the feedback reports. However, management objected to including detailed information per type of copier in the reports, because this would take too much effort to program and maintain. As a consequence, some improvising would become necessary in order to get information per type of copier to be used in discussion in small 'type groups'.

Results

This final stage in the development of ProMES resulted in a three-page individual feedback report and a two-page region feedback report (an example, translated from Dutch, is given in Appendix B).

During the operationalization of the indicators, it became clear that there would be large fluctuations in individual monthly indicator values (especially for percent repeat calls and percent return calls due to lack of car stock). The design team felt that feedback covering a larger period would be necessary to detect whether an indicator is really improving. A six month period was considered long enough to get a more stable measure. Using a *moving average* of six months also increases the reliability of the indicator values of the indicators measured by the supervisor of which it was doubtful whether the monthly sample would be large enough.

To be able to compare productivity of technicians, the monthly overall effectiveness score is expressed as a *percentage of maximum effectiveness*. This index is a measure of how well the technician is doing relative to how well he could be doing. This additional measure had become necessary, because the possibility existed that data on some indicators were not available in a certain month. For example, a technician may not have had the opportunity to claim spare parts. In this case the indicator 'Completeness of claims' does not apply to him and therefore neither an indicator value nor an effectiveness score can be obtained; if the maximum effectiveness score for 'Completeness of claims' is excluded from the maximum overall effectiveness score, then the percentage of maximum effectiveness is a better indicator for overall productivity than the absolute overall effectiveness score.

Individual feedback report (Appendix B)

The first page of the report showed how the *total effectiveness score* was composed of the effectiveness scores of products and indicators, for the last month and for the last half year.

The *type-dependent indicators* were described on the second page of the report. The total effectiveness score for an indicator was the weighted average of the effectiveness scores calculated for the different types of copiers. The weight was the number of calls the technician performs on each type, indicating the relative importance of each type of copier in that particular month. Only indicator values based on five or more calls on a type of copier were included. An example of this weighting procedure is given in Figure 4.4.

Indicator: Mean Copies Between Calls			
type of copier	number of calls	indicator value	effectiveness score
A	24	12.000	17
B	6	13.700	−23
C	11	26.400	35
D	2		
E	15	19.200	−2
Total effectiveness MCBC (= weighted average)			11

$$\text{Weighted average MCBC} = \frac{(17) * 24 + (-23) * 6 + (35) * 11 + (-2) * 15}{24 + 6 + 11 + 15} = 11$$

Figure 4.4. Type dependent indicators: weighted average procedure for calculating effectiveness values.

As explained, some of the MCBC results originated from previous months, but were attributed to the month in which they could be calculated. Therefore, the calls used as weights in calculating the effectiveness score on MCBC were not the same as the ones used as weights for the other type-dependent indicators.

The moving average effectiveness scores were not directly linked to the moving average indicator values, because the contingencies were based on monthly indicator values. It would be incorrect to use these contingencies to calculate the effectiveness scores of indicator values averaged across a 6 month period, which have a smaller variance.

The composition of the effectiveness scores on the type-independent indicators was shown on the third page of the feedback report.

Region feedback report (Appendix B)

The region feedback report was an aggregate of the feedback reports of all individual technicians in the region. All technicians contributed equally to the region report[4]. The region effectiveness score on each indicator was the average of the individual effectiveness scores on that indicator (except, of course, for the group level indicators). The

[4] Eventually, this was changed into a weighted contribution, which means that technicians who received feedback on a large part of their work contributed more to the region feedback report than technicians who received feedback on a small amount of work. The monthly number of hours of labor time covered in the individual report was used as the weight.

structure of the region report was identical to pages one and three of the individual report.

Programming of feedback reports

The technical implementation of the system was performed by a system analyst from the data processing department, based on detailed specifications drawn up by the facilitators. This technical implementation included programming of feedback reports and input and output facilities and took about five weeks. The facilitators entered the contingencies into the system. The result of the technical implementation was a program that generated (per month, per region) a set of individual feedback reports and a region feedback report. Besides the feedback reports, the program generated indicator data per type of copier and a frequency distribution of indicator data, both covering a one year period. These were used by the facilitator for establishing the contingencies. The program included input facilities for entering the contingencies, region indicator scores, and types of copiers to be excluded from the system.

4.4.5. Review and approval meetings

The formal ProMES approach contains two review and approval meetings in which the design team presents parts of the system to the management and consensus is reached on the final contents of these parts (Pritchard, 1990). Typically, these meetings are organized after the indicators have been developed and after the contingencies have been established.

For two reasons, the review and approval process in the Nashuatec project was more complex. The first reason was that the system was developed in two regions that did not make the same progress (for example, the Enschede region had already established the contingencies while the Utrecht region was still working on the indicators). Secondly, the design process in the two regions two regions was to result in *one* system applicable to the whole field service department. Although these issues did not influence the review and approval meetings on the indicators (which took place in both regions separately), the process of getting final approval was more complicated. First, in both regions, a separate meeting took place to achieve *consensus within regions* on the products, indicators and contingencies (a tentative agreement on the system between Enschede and the management had been reached, but new issues had come up during the delayed design process in Utrecht). Both regions' preferences were confronted during an *inter-region meeting*. During this meeting, representatives from both regions achieved consensus on the proposal to be discussed with management during the final review and approval meeting. The review and approval meetings will not be discussed as such; salient issues that came up during these meetings have been mentioned in the previous sections.

4.4.6. Overview of the design process

Design problems and solutions

Several design problems have been taken into account in designing the ProMES system. An overview of these problems and the corresponding solutions (at least on paper, since the use of the 'prototype' in the feedback meetings will show to what extent these solutions have been effective ones) is given in Table 4.6.

Table 4.6. Design problems and solutions.

Design problem	'Solution'
"Interdependencies" (in the context of an individual task)	- individual indicators (disentangling of contribution of individual technicians to region performance); - some indicators stimulate cooperation (preventive maintenance, History Card); - combination of individual and group feedback (both accountability and stimulation of cooperative behavior).
"Complexity of work flow" (different types of copiers)	- type-dependent indicators (and contingencies); - procedure for establishing large number of contingencies.
"Lack of communication" (communication between design regions)	- facilitators as messengers; - inter-region meeting.

In Table 4.7, data are presented on the number of meetings used to design the system in Enschede and Utrecht. It took 16 months to develop the complete ProMES system, from June 1989 until September 1990. This period included two holiday periods during which no meeting were scheduled. During this period, two students each spent, in succession, nine months full time on the project to fulfil their master thesis' obligations. Their job was to prepare the meetings, to act as (co-)facilitators, to assist in adapting the company's computerized information system to produce ProMES reports, to do most of the paperwork involved (reports, questionnaires, and so on) (Coolen, 1990; Kleingeld, 1990).

In both regions, there have been 13 design meetings (of 2½ to 3 hours each), including two review and approval meetings with management and one meeting to reach consensus between the regions. Developing the indicators turned out to be the most time-consuming activity, taking 2½ in Enschede and 3½ meetings in Utrecht. The development of contingencies took 3 meetings in Enschede and 2 meetings in Utrecht. Latter numbers would have been much higher, if the scaling techniques had not been used. As is usually the case, developing products did not take up a lot of time (1 meeting in Enschede and 1½ meetings in Utrecht). There were separate review and approval meetings with management in each region on the products and indicators, and also on the contingencies. Before the second review and approval meeting, there was one meeting in both regions to achieve final consensus on the system, and one inter-region meeting to prepare for the second review and approval meeting with management. In the last development meeting, the design team developed the feedback report and discussed ways of handling the feedback meeting. In both regions, 33 hours were spent on the development meetings.

Table 4.7. Overview of the design process.

Design phase	Enschede	Utrecht	Pritchard (1990)
Introduction	½	1	1
Developing products	1	1½	2
Developing indicators	2½	3½	5-8
Review and approval management (products & indicators)	1	1	1
Developing contingencies	3	2	3-4
Review and approval management Enschede (contingencies, tentative)	1	-	
Consensus within regions	1	1	
Consensus between regions	1		
Final review and approval management (Utrecht: whole group; Enschede: representatives)	1		1
Developing feedback report and feedback meetings	1	1	2
Number of design meetings	11+2	11+2	15-19
Time spent per meeting	2½-3 hrs	2½-3 hrs	1½-2 hrs
Time spent overall	33 hours	33 hours	26-33 hours
Total development time: June 1989 - September 1990: 16 months			7-9 months

The time invested per design phase is comparable to the rough estimate given by Pritchard (1990), if one takes into account the fact that the meetings for achieving consensus within and between the two design regions mainly focused on the indicators. Although the equal number of meetings in Enschede and Utrecht suggests that the design process progressed equally well in both regions, this is not the case. Because of the resistance encountered in region Utrecht in the start-up phase of the design process, the 'cooperative' region (Enschede) was a few months ahead. This means that most of the problems mentioned in this chapter (controllability of indicators, differences between types of copiers, measuring long term quality, etc.) were encountered first in this region. By the time Utrecht 'arrived at' these design problems, the facilitators had gained a better understanding of the specifics of a technician's job, the measurement problems and possible solutions. As a result, the process became less difficult and less time-consuming than might have been the case had Utrecht been the first or only region taking part in the design of ProMES.

Costs (time and expenses)

Tables 4.8 and 4.9 give an overview of the costs involved in designing the system. Data is presented on time invested by the parties involved in different categories of activities (Table 4.8) and direct expenses related to the design process (Table 4.9).

Table 4.8. Design process: time spent.

Time spent (estimate, hours)	design meetings	between meetings	steering committee	writing program	overall
Technicians	1000	50			1050
Supervisors	100	100			200
Management	30	50	100		180
System analyst				200	200
Master's student (1½)	150	1100	30		1280
PhD student (½)	30	500	10		540
Assistant professor (1)	130	150	40		320
All parties combined	1440	1950	180	200	3770

Table 4.9.

Design process: direct expenses.

(Dfl 1.00 = approx. $ 0.55)

Category	Expenses (estimates)
Design meetings	
payment of technicians/supervisors	Dfl 25,000
meeting rooms/facilities/dinner	Dfl 28,000
Facilitators	
Salary and expenses	Dfl 43,000
Overall expenses	Dfl 96,000

4.5. IMPLEMENTATION OF THE SYSTEM

4.5.1. Pilot feedback meetings

At the start of the implementation, the idea was to incorporate the feedback meetings into the monthly region meetings that were conducted by the supervisor. However, it was decided to hold the first feedback meetings separately, because they would take up considerable time (about 2½ hours). They were conducted in part by the facilitators, in part by the supervisor. The five pilot feedback meetings will be discussed.

Each month, data on the individual indicators and group indicators were collected, and feedback reports were generated for all technicians in both regions. The feedback meetings took place between the fourth and ninth work-day of each month, to ensure timeliness of the feedback. It was decided to distribute the feedback reports during the feedback meetings. After each feedback meeting a memorandum was sent to all participants, confirming decisions and points for attention.

4.5.2. Procedure during pilot feedback meetings

The pilot feedback meetings were carried out according to the subdivision agreed upon (see Table 4.5). First, plenary discussion of the three group indicators and the two type-independent individual indicators; then, small-group discussion on the five type-dependent individual indicators; finally, plenary announcement of the outcome of the small-group discussions.

Group indicators

Clear agreements on the group indicators ('Compliance with maintenance procedures', 'Accuracy of History Card' and 'Correctness of behavior') to which all technicians should adhere are a precondition for the validity of many of the individual indicators. The supervisor's observations during the past month pertaining to these indicators were discussed with the entire group. The supervisor explained how the effectiveness score on each indicator came about–using the contingencies–and compared this score with effectiveness scores of previous months. He illustrated his evaluation with (anonymous) examples of positive and negative incidents he witnessed in the past month, some of which were shown on transparencies. Because of the routine nature of most elements of which the indicators consist, discussion was usually limited to directing attention to one or more areas that might be improved and getting group commitment on improving them.

Type-independent individual indicators

In the initial discussions on the indicators 'Percent capacity used' and 'Completeness of claims', it became evident that there were major problems with the accuracy of the data.

In the case of 'Percent capacity used', the facilitators tried to solve the problem by introducing the system to the planners in the central planning department who are responsible for entering data on the technicians' activities into the information system. As a result, the planners decided to check on missing data at the end of each month. Meanwhile, a number of technicians kept a written log on all time spent in order to check on the ProMES data. This resulted in accurate data by the fifth pilot feedback meeting. However, even after such efforts, this indicator still remained very susceptible to errors, which can be very detrimental for scores at the individual level.

'Completeness of claims' turned out to be a hopeless case, primarily due to the complexities of the claiming procedure used by the department handling the financial aspects of claiming, which prevented accurate feedback to individual technicians.

In all, these problems took up a lot of time during feedback meetings and thus left little time for discussing the type-dependent indicators, which are considered the most important part of the system.

Type-dependent individual indicators

The part of the pilot meetings which dealt with the type-dependent individual indicators (MCBC, percent repeat calls, parts cost per call, labor time per call) contained three major activities, which will be discussed: individual interpretation of the feedback reports, visual representation of effectiveness scores, and discussion in small groups.

During the third feedback meeting, the technicians were asked to individually interpret their own feedback reports using the contingencies and a description of the calculations involved. A number of questions was asked:

- Which aspects of the feedback report are not clear yet?
- What does the feedback report tell you about your strengths and weaknesses and about the strategies you follow?
- Are you satisfied with your effectiveness score? If not, what strategies could you try out to improve your effectiveness?

According to expectations, the lack of participation of the technicians in the operationalization of the indicators and the complexity of this part of the report lead to a number of questions that could be answered by the facilitators, either immediately or after checking with the programmer. There was, however, one exception: percent return calls due to lack of parts in car stock. The objective of this cost indicator (measuring to what extent a technician avoids return calls by maintaining a sufficient level of car stock) could not be achieved. Technicians are allowed to order car stock parts together with non-car stock parts. For this reason, it cannot be determined whether the car stock part was ordered to finish the call or as a normal resupply of the car stock, thereby preventing accurate measurement of a technician's supply control.

Although most technicians agreed that the feedback reports were useful tools to monitor and improve performance, the reports were considered complex and difficult to interpret, especially for the type-dependent indicators. For that reason, visual representation of effectiveness scores was added to the feedback reports. This consisted of line graphs representing individual and region effectiveness score over time for the type-dependent indicators. Using these graphs, the technicians could see at a single glance 'where they stood', and whether they had succeeded in improving both in absolute terms (compared to zero point) and in relative terms (compared to the region as a whole). Unfortunately, it was not possible to have the computer system generate the graphs. Therefore, the technicians had to fill in these graphs themselves. Whereas some technicians found this to be useful, others had serious reservations about this 'clerical' exercise. Although these graphs were just very simple, a lot of technicians for the first time got an idea about which–often implicit–strategies they had been using. For example, some technicians appear to have a 'high quality–high cost' strategy, whereas others appear to have a 'low quality–low cost' strategy. Because quality and cost are about equally important in the system, both strategies could in principle be equally effective. By comparing their graphs with colleagues who have approximately the same models, strengths and weaknesses of individual technicians become apparent (for example, technician A uses less labor time than his colleagues B and C on the same type and attains approximately the same quality-scores).

The objective of small group discussion was to learn about causes of performance differences between technicians by discussing individual performance data of those who work on the same types of copier. The focus should be on how to learn from each other and how to work more effectively and efficiently by using better strategies (so, for example, find out how technician A succeeded in limiting his labor time spent and assess whether technicians B and C might attain a higher score by following his strategy). Only during the fifth feedback meeting, was an attempt made to discuss these types of issues. This was only partially successful, because some technicians did not like to talk about their individual performance. Yet, this had become necessary because of the lack of information per type of copier: technicians had to exchange a lot of information to get to know the extent of individual differences.

In summary, the five pilot feedback meetings were only partially successful. The part of the meetings that focussed on the group indicators progressed satisfactorily. Yet, the recurring discussions on the validity of some of the indicators and lack of insight into the feedback reports hindered effective discussions on how to improve productivity on the individual indicators. The productivity effects obtained during the pilot feedback period are discussed in the next section.

4.6. EFFECTS ON PRODUCTIVITY

4.6.1. Design

Effects of the system were evaluated using a time series design with two experimental units (Enschede and Utrecht) and nine control units (the nine other regions that existed at this stage of the project). In both experimental regions there was only one month of baseline data (October 1990). In the experimental regions the system was implemented by means of monthly feedback meetings, during which there was feedback, but no specific goal setting. In the control regions, data on all indicators were collected but no information was fed back. The supervisors of these control regions were to collect the data on the group level indicators in order to get complete control group data. They had been given a brief overview of the ProMES system used in Enschede and Utrecht and a thorough instruction on the use of the group indicators. They received explicit instructions not to feed back any more information on these indicators than they had done in the past.

The period under consideration is November 1990 through May 1991. Feedback meetings in Enschede and Utrecht took place from November until March 1991. The meetings in April and May were used to evaluate the system. These will be discussed in Sections 4.7 and 4.8, as well as results from June 1991 through March 1992.

4.6.2. Group indicators

The group indicator data are shown in Figure 4.5. In the ProMES regions (Enschede and Utrecht) the total effectiveness score on the group level indicators improved remarkably after the first feedback meeting (from –33 to 25). In the following months, the effectiveness score dropped a little and stabilized around 15. Overall, the mean effectiveness of the ProMES groups during the feedback period was 15 and the overall mean for the control groups was –25.

The group effectiveness scores of Enschede and Utrecht during feedback are significantly higher than the effectiveness scores of the control regions (see Table 4.10). The small increase in the control regions may have been caused by the fact that collecting data on the group indicators supplied the supervisors with useful information.

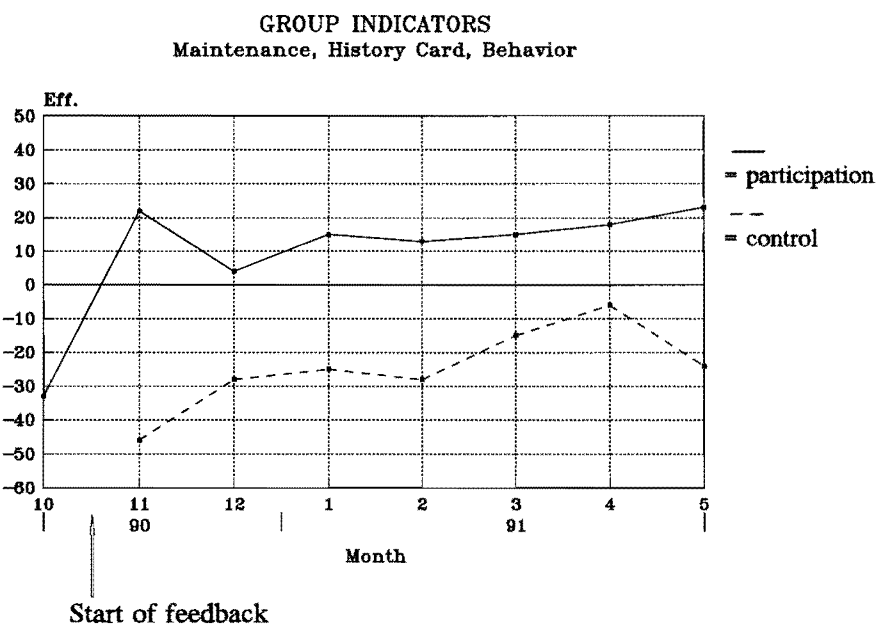


Figure 4.5. Group indicators: effects on productivity (pilot feedback period).

It should be noted that the feedback of information on group level indicators has had an immediate positive effect on productivity in Utrecht and a smaller and delayed effect in Enschede. Two reasons for this occurrence can be given. First, during the first feedback meeting, region Utrecht was confronted with extremely negative group effectiveness scores (an overall effectiveness of −75). The perceived need for improvement was thus higher than in the Enschede region, which started around zero effectiveness. Secondly, Utrecht's supervisor listed points for attention in a 'tell and sell' way, which is considered effective (e.g. Latham, Erez & Locke, 1988; Latham & Saari, 1979a; Latham & Wexley, 1981), whereas the supervisor from Enschede seemed less able to communicate the results clearly, especially during the first few feedback meetings.

As noted in Section 4.4.1, a satisfactory group indicator effectiveness is a prerequisite for valid use of the individual indicators.

Table 4.10. Comparison of group level effectiveness (pilot feedback period)

	Enschede/Utrecht	Control regions	Paired t-test (df=6)
Preventive maintenance	4	−25	P = .001
History Card	−2	−7	P = .029
Behavior	13	7	P = .006
Overall group effectiveness	15	−25	P = .000

Accuracy of measurement

The group indicator scores are based on the judgement of the region supervisors. To investigate whether differences between regions/experimental conditions are actual differences in performance or just differences in the supervisors' interpretation of the checklist elements, an 'inter-rater reliability test' was done. For this purpose, a technician performed repairs and maintenance on a photocopier used at the company's head office. Purposely, some elements of the maintenance were not carried out according to the specifications of the preventive maintenance procedure and some mistakes were made in the completion of the History Card. The next day, all supervisors then available (9) were asked to inspect the photocopier and fill in the checklists for 'Compliance with preventive maintenance procedure (after visit)' and 'Accuracy of History Card'. The results showed a rather large variance in the judgements made (preventive maintenance: 60–100 percent; History Card: 50–80 percent). Differences on the individual checklist items were discussed and consensus was reached on the interpretation of each item. Also, some items were changed, added to or removed from the checklists.

The test-results showed that the two supervisors in the ProMES condition arrived at the same mean scores as their colleagues from the control condition, both on the Preventive maintenance and the History Card-indicator (with the supervisor from Utrecht being more critical than his colleague from Enschede). This limited test suggests that the positive results obtained in the ProMES condition on the group indicators reflect actual improvements rather than more lenient measurement by the supervisors.

4.6.3. Individual indicators

Type-dependent indicators[5]

The experimental regions show no change in productivity compared to the control regions during the first seven months of feedback[6] (see Figure 4.6). During the first seven months of feedback, the mean effectiveness of the experimental groups was 16 and the overall mean for the control groups was 19, a negative difference of 3 points (see Table 4.11). A t-test on the mean overall effectiveness scores of 183 technicians (32 experimental, 151 control technicians) revealed that this difference was not significant ($t(181)=-.70$; $P=.490$). The experimental regions attained approximately the same effectiveness score (Enschede: 14, Utrecht: 17).

Table 4.11. Individual effectiveness scores in the experimental and control regions.

	Enschede	Utrecht	Enschede/Utrecht	Control
MCBC	0	5	2	–1
Percent repeat calls	11	10	11	11
Parts cost per call	–5	–5	–5	0
Labor time per call	8	7	8	9
Overall effectiveness	14	17	16	19

A reconstruction of baseline results (taken from the moving average data of the first feedback report) revealed that in the six month period before feedback was given, the average effectiveness of Enschede/ Utrecht was 30 (Enschede: 31, Utrecht: 29), whereas the average effectiveness of the control regions was 34, a negative difference of 4 points. Thus, the difference between both groups remained the same when comparing the baseline period with the pilot feedback period. The absence of positive effects may in part be explained by the fact that relatively little attention had been paid

[5] Excluding 'Percent return calls due to car stock', because of difficulties in obtaining valid measures. Evaluation of effects includes those individual indicators that were considered valid during the entire pilot feedback period. These are 'Mean Copies Between Calls', 'Percent repeat calls', 'Parts cost per call', and 'Labor time per call'.

[6] The region effectiveness score is the average of the technicians' effectiveness scores. The ProMES system does not include fax machines, laser printers, and some types of photocopiers. The machines not included in ProMES represent only a small percentage of the total workload of a region. However, some technicians who work primarily on these machines only receive feedback on a small portion of their total workload. The results of these technicians (both in Enschede/ Utrecht and in the control regions) are not included in this evaluation of effects. Only those technicians who spend at least 25 percent of their labor time on copiers measured in ProMES in a month are included.

to these indicators during the feedback meetings. Intensive group discussion, which was considered necessary to arrive at better strategies to improve productivity, hardly took place.

In Section 4.7, a thorough analysis of the causes for the lack of productivity improvement is described.

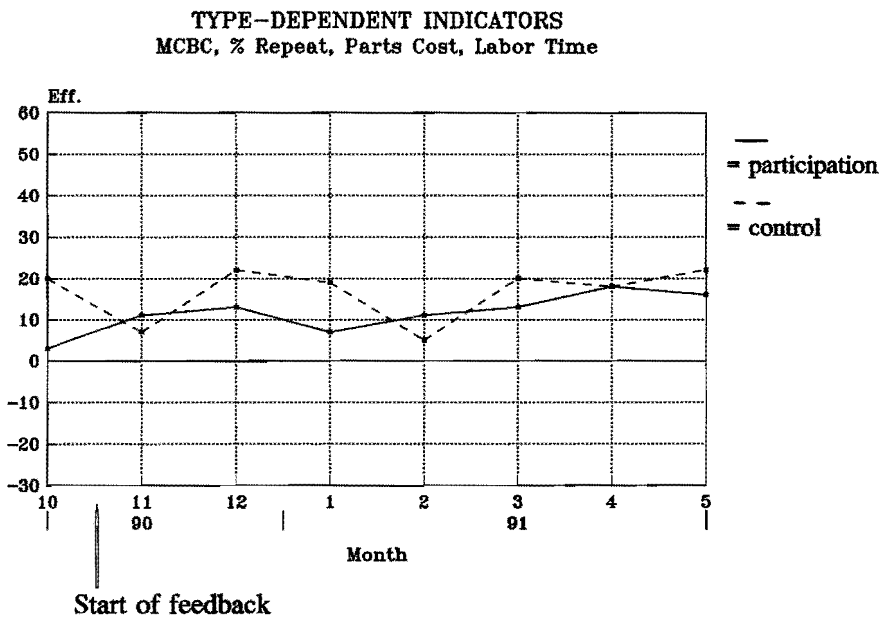


Figure 4.6. Type-dependent indicators: effects on productivity (pilot feedback period).

Type independent indicators

Due to difficulties in obtaining valid data on 'Percent capacity used' and 'Completeness of claims', no data are available on these measures. However, there is anecdotal evidence of some improvement in Enschede and Utrecht. According to the planning department, after feedback started technicians from Enschede and Utrecht were concerned about attaining 100 percent use of capacity: they would ask their dispatcher for an extra call if they felt they might not be able to 'make their hours'. Also, during the feedback meetings it became clear that the technicians had been paying special attention to claims in order to check the validity of data that is fed back.

Summary

On the basis of these results, the main conclusion must be that implementation of the ProMES system (feedback of measurement information) has not produced the effects hoped for. Although positive effects have been obtained on the group indicators, no change occurred regarding the type-dependent individual indicators, which are much more important. In other words, although design problems posed by several situational characteristics of the Nashuatec setting have been accounted for in the design process and in the design, a pilot-implementation of the design (the 'simulation' phase in the design cycle) was only partly successful. A logical next step within this design-approach is a comparison of the design to the desired characteristics to find the causes for the lack of positive results and generate solutions to eliminate these causes (e.g. by redesigning the system or changing the context). This is the topic of the next section.

4.7. PARTICIPATIVE PROBLEM ANALYSIS

Although some improvements were made in group level effectiveness, all parties concerned (management, design regions and facilitators) were disappointed with the overall effects of the feedback provided by the ProMES on the performance of the technicians who had developed the system. To deal with this problem, a participative problem analysis was done, which resulted in six perceived main causes for the problem and six corresponding solutions.

4.7.1. Process

First, the basic issue of potential performance improvement was settled: technicians were of the opinion that, notwithstanding the effects of performance-improving measures taken at earlier times by the management, there still was room for additional improvement. The management, when asked so informally, was of the same opinion. Although they could not give indications as to the degree of potential performance increase, they could give anecdotal evidence indicating that at least some efficiency improvement had to be possible.

After that, a participative problem analysis was carried out in order to gain insight into the possible causes of the lack of productivity improvement. This analysis consisted of four steps. First, possible causes were generated by means of interviews with the region supervisors, and brainstorming sessions and small group discussions with the technicians in the design teams. Using checklists compiled by the facilitators, the possible causes (19 in all) were then rated by the individual technicians. From this procedure, six main perceived causes emerged. After that, the design teams, together with the facilitators, developed solutions for these causes. Finally, the proposed solutions were presented to management.

4.7.2. Perceived causes and proposed solutions

The six main perceived causes and proposed solutions are displayed in Table 4.12.

The first perceived cause relates to the role of management in the ProMES program. The design teams doubted whether management was still committed to the ProMES program. The technicians felt somewhat abandoned since they had been struggling for a number of months with the feedback meetings without getting regular signals from the management that this was still appreciated and in conformance with management policies and intentions: "Is this still serious, or do they have other interests and priorities right now? If so, why should we bother?". Thus, a clear management statement on the importance of the project was asked for.

Secondly, technicians asked for the removal of three invalid indicators from the feedback reports. A lot of discussion time during feedback meetings had been spent on validity issues. On a number of occasions, technicians successfully demonstrated that data in the feedback reports were incorrect. This resulted in feelings of distrust which did not stimulate performance improvement initiatives. Besides, not much time was left in the meetings for discussing such initiatives. The indicators involved (percent return calls car stock, percent claims, and percent capacity used) were the three least important indicators[7]. It was felt that the completeness of the system would not be compromised by removing these indicators.

Thirdly, the feedback report was considered too complex. Technicians got overall ProMES effectiveness scores on each indicator. In addition, per indicator, for each separate type of photo copier, both indicator scores and ProMES effectiveness scores were presented. The addition of graphical presentation of this information was deemed necessary to be able to assess one's productivity at one glance and to be able to discern trends quickly. This graphical information was not readily available.

The fourth cause concerned the division of the feedback report. In the feedback report, group effectiveness scores on the indicators measured by sampling were included in the individual feedback reports of each technician, thereby suggesting that the scores represented the individual performance of the technician during the period concerned. Technicians refused to be held personally responsible for the performance of their colleagues and asked for removal of these indicators from their individual reports. The solution proposed was to present a feedback report containing overall region scores on all valid indicators (including those measured only by sampling) to the whole group, and to present individual feedback reports containing only information on the valid individual indicators to each individual technician.

The fifth cause pertained to the procedure used during the feedback meetings. The design teams were of the opinion that discussion in small groups would not constitute the best way of working with ProMES. In the few instances when these discussions were attempted, some difficulties had become apparent. First, some technicians had not been willing to share information about their performance with their colleagues. At the same time, senior technicians, who were the 'primus inter pares' in these groups sometimes had lower effectiveness scores than their less experienced colleagues, whom they were supposed to help (e.g. because of the higher degree of difficulty of the calls the seniors had to perform). This seeming contradiction did not facilitate the discussions.

---

[7] Results of a cost-calculation, reported in section 4.7.6, reveal that percent return calls car stock, percent claims, and percent capacity used had an importance compared to the most important indicator (MCBC) of 16%, 1%, and 22%, respectively.

Table 4.12. Main perceived causes of the lack of productivity improvement, solutions, and contribution of these solutions to achieving an 'accepted control loop'.

perceived cause	solution	contribution to 'accepted control loop'
1. No visible management commitment	management statement about: - objectives - priority of ProMES - willingness to support the program	*Context factors:* - visible management commitment - trust between incumbents, supervisor, and management
2. Invalidity of three indicators	Removal of invalid indicators from feedback report (10 percent of total productivity)	*Feedback criteria:* - (perceived) validity - completeness
3. Complexity of feedback report	Addition of graphical feedback to the report	*Feedback criteria:* - understandability
4. Division of feedback report	Excluding group indicators from individual feedback report Division of group feedback report into separate individual and group sections	*Feedback criteria:* - understandability (- validity)
5. No positive consequences of productivity improvement; unclear connection with performance appraisal	Use of ProMES as one of the main elements of the performance appraisal system (to determine the year end bonus)	*Feedback criteria:* - perceived importance *Context factors:* - consistency with reward systems
6. Small group feedback meetings unsatisfactory	Individual indicators: bilateral feedback meeting (individual technician and supervisor) Group indicators: group feedback meetings (whole region and supervisor)	*Goal criteria:* - challenging (difficult & attainable) - commitment as a result of problem diagnosis and choice of improvement strategies

Lastly, the small groups were composed of technicians working on the same types of copiers. Some technicians, however, did not fit into these groups, because they were the only ones working on an unusual range of types. Because a large part of the system focuses on the work of individual technicians, it was proposed to discuss the individual feedback reports in bilateral meetings of an individual technician and his supervisor. The region reports, including the group indicators, could be discussed in a group feedback meeting.

A sixth cause played a role, but did not come explicitly to the surface until the management reacted to the first five solutions proposed by the design teams. It had to do with the issue of "what is in it for us, if we succeed in improving our performance?". In fact, opinions on this point were mixed, ranging from "you are already paid for doing a good job, so why should you get more for eventually really doing a good job" to "if we improve, the company earns a lot more money, why shouldn't we get our share; besides, it stimulates to have an outlook on a reward". A consideration in this respect was the availability of individual productivity data to the supervisor, which would render a performance appraisal without (implicit) use of these data nearly impossible.

4.7.3. Discussion with management

The proposed solutions were brought to the attention of the management. Management agreed to arrange a meeting at which they would make a statement and react to the concrete suggestions of the design teams. The management statement at these meetings came as a surprise. In short, management took the position that, although ProMES had not brought the improvements expected, it was still very worthwhile to keep it and use it for performance appraisal purposes. This statement caused a lot of confusion, given the guarantee at the start of the project not to do such a thing unless all parties concerned would agree (see Section 4.2.3). In a long discussion, the proposed solutions were combined into a plan acceptable for all. To begin with, although the technicians were outraged at first, they soon understood that the management proposal fitted in very nicely with their sixth problem, the absence of a clear link between performance improvement and reward. It was agreed that their annual bonus would be linked to their performance as measured by ProMES. This would be done by making ProMES an integral part of the performance appraisal system, by means of which the amount of annual bonus is determined. The exact procedure would be worked out and presented to the technicians for approval. A precondition was that only valid information would be used, so the discussion continued on the other solutions proposed by the technicians. In fact, all suggestions made were accepted by the management. In short, agreement was reached on the content of the individual feedback reports: only the indicators agreed to be valid were included. These were 'Mean Copies Between Calls', 'Percentage repeat calls', 'Parts cost per call', and 'Labor time per call'. Thus, only the two most important products, quality and cost, were officially included in the individual reports. The region report would contain information on the overall score of the region on those four individual indicators, and in addition to that the scores on

'Compliance with preventive maintenance procedures', 'Accuracy of History Card' and 'Correctness of behavior', the three indicators measured by a sample taken by the supervisor. The request for graphic information presentation was granted too. Finally, it was decided to have regular meetings with the management in the future to evaluate the progress on the ProMES project and to discuss mutual affairs.

During the following two months (July and August 1991), the proposed changes were implemented with the exception of the 'bilateral feedback meetings', since this new approach necessitated thorough preparation, including a feedback training program for the region supervisors (which will be discussed in Chapter 5). Feedback resumed in September 1991 (feedback reports covering the previous three months were supplied in the first week of September). Until the start of the bilateral feedback meetings (April 1992), the technician received their feedback report by mail; no 'official' feedback meetings took place.

4.7.4. An accepted control loop?

The results of the potential problem analysis, as represented in Table 4.12 (third column) indicate that most of the factors inhibiting substantial productivity improvement relate to criteria for an accepted control which were not met. As discussed in Chapter 2, the essential elements of a control loop are goals and feedback. ProMES can be considered a method for designing and implementing accepted control loops (see Section 2.4 and Kleingeld & Van Tuijl, 1992, in press). A ProMES system can only be effective to the degree that the feedback provided by the system that is developed, as well as the goals implied by the system, are accepted by those who are supposed to regulate their own performance by means of the system. Only discrepancies between accepted goals and accepted feedback will result in attempts to reduce these discrepancies by investing effort in strategies leading to high performance. The model of the accepted control loop helps us to explain why the pilot implementation of the ProMES design did not have overall positive effects on the productivity of the technicians involved.

For several reasons, feedback acceptance might have been low. As discussed above, the feedback reports were rather complex due to the necessity to take differences between types of copiers into account. The reports provided indicator data for each type of photocopier serviced by a technician and combined these data through the contingencies that belonged to each type into overall effectiveness scores for each indicator. In addition, six month moving averages were provided. Feedback was given as 'raw' indicator data, or in the form of effectiveness scores. For most indicators, information referred to the prior month. The MCBC indicator, however, referred to an undefined period in the past. For these reasons, most technicians had difficulty understanding the system, notwithstanding the fact that they had assisted in developing it themselves. It will, undoubtedly, be difficult to accept and use feedback that one does not understand completely. Acceptance will also have been lowered by recurring discussions on the validity of some indicators and the technicians' unwillingness to accept individual responsibility for the group indicators.

For accepted discrepancies between feedback and goals to occur, in addition to accepted feedback, accepted goals are a prerequisite. It is doubtful whether there were accepted goals in this setting. This is caused by the way in which the contingencies were developed. In estimating maxima, zero points and minima, the performance data from all (more than 200) of the company's technicians working all over the Netherlands were used. The zero point was determined by calculating the average indicator value of all technicians. The minimum and maximum indicator values were, respectively, the worst and best indicator values achieved by a technician in the measurement period. By implication, many technicians belonged in the lower half of this distribution, and many belonged in the upper half. The former group will receive mostly negative effectiveness scores, the latter group mostly positive ones. For this reason, a lot of technicians will have difficulty in accepting the maxima, zero points, and minima as relevant for themselves as individuals. As a consequence, they will have difficulty in defining relevant performance ranges for themselves and assessing which effectiveness level would represent a difficult but attainable goal. In other words, instead of helping them to establish specific, and difficult but attainable goals, the contingencies might very well confuse technicians about the goals to be strived for. In summary, in designing the system, individual differences between technicians have not been taken into account sufficiently. This problem cannot be solved by changing the contingencies, since this would further increase the already high complexity of the system. Rather, the feedback procedure should contain ways for dealing with these individual differences which should facilitate the setting of individual goals.

Accepted discrepancies between feedback and goals should result in the decision to invest effort in a strategy that leads to a maximum performance increase in the future. In this case, however, deciding for such a strategy is a complicated affair, because the contingencies of all the types of copiers the technician is working on have to be taken into account. In fact, there are several options. The technician could choose to improve his performance on the type of copier on which improvement on all the indicators would result in the largest increase of his overall effectiveness score. The technician could also decide to improve his performance on the indicator on which an overall improvement on all the copiers would result in the largest increase of his overall effectiveness score. Of course he could decide to use some combination of the two options mentioned. Because of the large number of variables involved, it is very difficult to determine the best strategy to follow (it is even hard to determine a good or acceptable one) (Kleingeld & Van Tuijl, 1992, in press).

Aside from the absence of necessary control loop features, two context factors seem to have contributed to the lack of positive effects. Firstly, there was no visible management commitment to the system. Although management had always been willing to pay the technicians for their work on ProMES during overtime hours, and had taken part in the review and approval meetings, it had adopted a rather instrumental attitude toward ProMES. They seemed to consider the system as a separate activity, which should be able to operate without additional support from the head office. The second factor was the lack of consistency between ProMES and the performance appraisal, which included a number of trait-like dimensions (e.g. flexibility, independence, task conception) and unspecified behavioral and result-oriented elements (communication with colleagues, quality of work, quantity of work) which, although not completely contrary to the ProMES measures, were not clearly related to them. This may have caused a rather low perceived importance of the ProMES feedback. In other words: the control-loop design has not been consistently linked to the context of other control system, which diminished its effectiveness. The link between ProMES and the performance appraisal system which (a result from the problem analysis) is shown in the next section.

4.7.5. The link between ProMES and performance appraisal

Procedure

After careful deliberation, management and design team agreed that ProMES should contribute to the distribution of the annual bonus (varying between zero and eight percent of a year's salary). 60 percent of this bonus would be determined by the supervisor's appraisal of the technicians performance, using the standard performance appraisal form. The remaining 40 percent would be determined by the ProMES scores on the four individual indicators. Half of this score would result from a comparison of the technician's average absolute effectiveness score in the appraisal period to that of all other technicians. The other half would be determined by the technician's relative effectiveness score (i.e. the difference between the period at hand and the period covered in the previous appraisal) compared to the relative score of all other technicians. See Figure 4.7 for a global outline of this procedure. A more detailed description of the procedure can be found in Section 6.5.

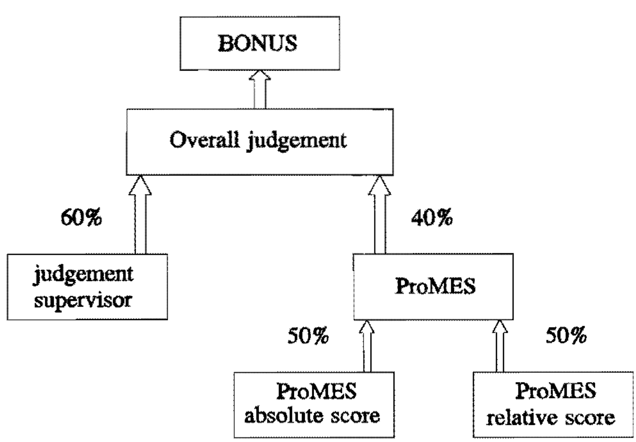


Figure 4.7. Contribution of ProMES to the annual bonus.

Evaluation

The use of ProMES information in the performance appraisal was evaluated by means of a questionnaire, distributed among the technicians of Enschede and Utrecht (December 1991). the main results are presented in Table 4.13. These show that the overall attitude towards the use of ProMES for performance appraisal purposes was positive. Although the technicians were in disagreement over whether the link between ProMES and the performance appraisal had been implemented overhastily, the majority considered it an improvement compared to previous appraisals and would like to use the same procedure the following year. They indicate that the connection with the performance appraisal caused a (small) increase in attention for the ProMES reports and that they would like to discuss their ProMES reports with their supervisor.

Table 4.13. Evaluation of the link between ProMES and performance appraisal (Mean scores and standard deviations).

Item [5-point Likert scale (1 = strongly disagree; 5 = strongly agree)]	Enschede n=19	Utrecht n=20	Overall n=39
A. Combining the supervisor's judgement with ProMES information results in a more accurate performance appraisal than a supervisor appraisal alone	4.26 (.99)	4.20 (.70)	4.23 (.84)
B. The use of ProMES information in the performance appraisal should have been postponed until next year	2.84 (1.53)	3.58 (1.07)	3.21 (1.36)
C. The link between ProMES and the ultimate performance appraisal is too complicated	2.68 (1.20)	2.60 (.82)	2.64 (1.01)
D. It would be a good idea to incorporate ProMES-information in next year's performance appraisal	4.16 (.96)	4.16 (1.12)	4.08 (1.03)
E. Because of the link between ProMES and the appraisal system, I have been paying more attention to the feedback reports	3.56 (1.04)	3.75 (1.13)	3.63 (1.08)
F. I would appreciate it if my supervisor were to discuss my ProMES results with me every few months during the next year	3.83 (.92)	4.25 (1.12)	4.05 (1.08)

For neither item, Enschede and Utrecht differ significantly at .05-level. Item B approaches significance (P=.095).

4.7.6. Re-establishing the contingencies: costs and customer satisfaction

When the relative importance of the indicators, as established in both design regions, was brought to the attention of management, an interesting situation occurred: management admitted that they were not able to adequately assess the accuracy of the groups' judgements. Although the choices made seemed plausible, management suggested a quantitative verification of the contingencies. Awaiting the results of this verification, the subjective estimates by the design regions were accepted by management as satisfactory estimates, which would be used in the pilot feedback period.

The *Activity Based Costing method (ABC)* (e.g. Cooper & Kaplan, 1988; Drury, 1989) was used to estimate the relative impact on cost reduction of comparable improvements on each of the indicators (Lamberts, 1991). Thus, for example, which cost reduction results from a 10 percent improvement on the indicator 'Parts cost per call' compared to a 10 percent improvement on 'Mean Copies Between Calls'? This analysis was carried out for three types of copiers, which represented the three product segments (low, middle, and high volume), and was then generalized to the whole population of photocopiers (see Appendix C for a brief description). In addition to this quantitative analysis, the relative *strategic effects* of improvements on the indicators were estimated in an interactive discussion between management and the design team. These strategic effects mainly included aspects of improved customer satisfaction and goodwill due, to, for example, shorter reaction times and less frequent malfunctions (especially within a short time period after a repair visit, less copier down-time, etc.).

The combination of cost-reduction-based and customer-satisfaction-based estimates of relative importance of indicators produced results that differed slightly from the subjective estimates that were made during the design process (see Table 4.14). MCBC remained the most important indicator. The importance of labor time per call increased, whereas the importance of percentage repeat calls and parts cost per call decreased slightly. Interestingly, the importance of 'Parts cost per call' is the result of this indicator's impact on cost reduction. This indicator does not contribute to customer satisfaction, since a large majority of the customers had a maintenance contract, in which replacement of spare parts is free of charge. Conversely, the importance of percentage repeat calls is mainly determined by its influence on customer satisfaction, through the reduction of faulty or insufficient repairs which compel a customer to again report a malfunction within a few days of the original repair.

Table 4.14. Comparison of the results of subjective estimates and cost reduction/ customer satisfaction estimates of relative indicator importance.

indicator	subjective estimate of design teams	cost reduction calculation (Act.B.Cost.)	contribution to customer satisfaction: subjective estimate	cost & customer satisfaction: combined assessment
MCBC	100	100	important	100
% Repeat calls	86	30	very important	70
Parts cost/call	61	82	unimportant	50
Labor time/call	67	96	somewhat important	85

4.8. EFFECTS ON PRODUCTIVITY

4.8.1. Individual indicators

Figure 4.8 shows the effectiveness scores of the two experimental regions and the control regions for the four type independent individual indicators. As reported in Section 4.6.2, the analysis of ProMES data from the initial feedback period (November 1990 through May 1991), which included the participative problem analysis (April and May 1991), revealed that no productivity increase had been obtained in the experimental regions compared to the control regions.

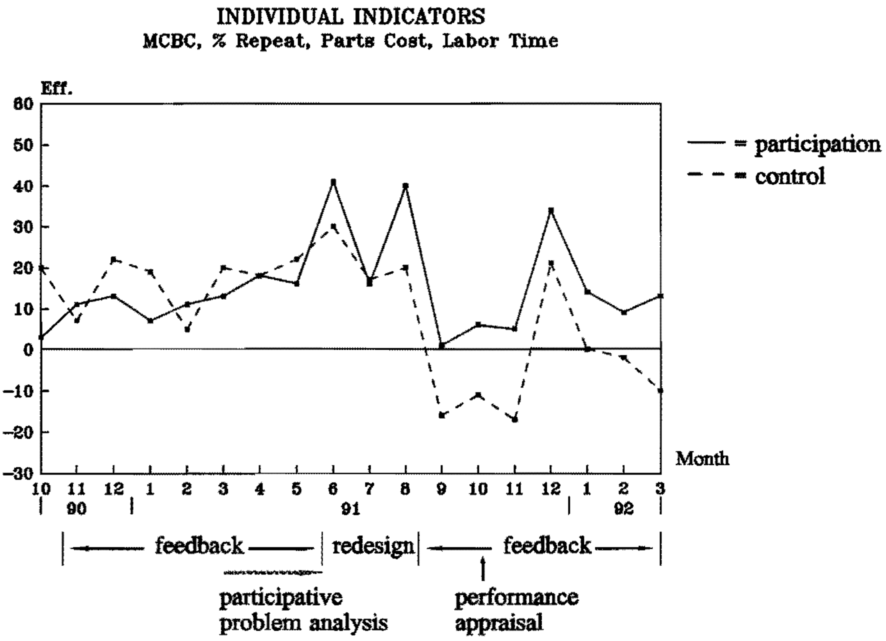


Figure 4.8. Individual indicators: effects on productivity.

In the 10 month period after the agreement with management had been reached (June 1991 through March 1992), the experimental regions achieved an effectiveness score consistently superior to the control groups (the overall mean effectiveness scores were 19 and 4 respectively, see Table 4.15).

Of the two experimental regions, the Enschede-region was somewhat more successful than the Utrecht-region: Enschede increased its effectiveness score from 14 to 22 (+8), whereas Utrecht's effectiveness score did not change (17).

Table 4.15. Comparing experimental and control group effectiveness scores.

period	pilot feedback (11.90-5.91)			post-redesign feedback (6.91-3.92)			change of Exp. relative to Control
indicator group	Exp	Con	Diff.	Exp	Con	Diff.	
MCBC	3	−1	4	2	−5	7	3
Percent repeat calls	11	11	0	6	5	1	1
Parts cost/call	−5	0	−5	2	2	0	5
Labor time/call	7	9	−2	9	2	7	9
Overall effectiveness	16	19	−3	19	4	15	18

A 2x2 Multivariate analysis of variance *(MANOVA)*[8] in which the mean overall effectiveness from the pilot feedback period and the post-redesign feedback period were considered repeated measures, revealed a significant interaction effect between the condition (ProMES or no ProMES) and time (before or after redesign of the system). The main time effect approaches significance (signifying an overall productivity decrease!), whereas the main condition effect is not significant. A planned comparison t-test on the mean change in effectiveness score of 183 technicians (32 experimental, 151 control technicians) revealed a significant difference between the experimental and the control condition. See Table 4.16 for detailed information.

Table 4.16. MANOVA and t-tests

MANOVA, repeated measures

	SS	DF	MS	F	sign.F
Between-subjects effects					
within cells	270163.39	181	1492.62		
condition (experim./control)	2012.27	1	2012.27	1.35	.124
Within-subjects effects					
within cells	82629.66	181	456.52		
time (before/after revision)	1440.83	1	1440.83	3.16	.077
condition by time	4222.70	1	4222.70	9.25	.002

Planned comparison of productivity change (t-test)

Condition	Number of cases	Mean	Standard Deviation	Standard Error
Enschede/Utrecht	32	3.7	30.10	5.32
Control	151	-14.2	30.24	2.46

t(181) = 3.04; P(one-tailed) = .002, significant

[8] Although the research design resembles a time series design, no time series analysis could be applied, because of an insufficient number of data points (Cook & Campbell, 1979).

Figure 4.9 gives a more simplified view of the productivity changes in the ProMES and control condition.

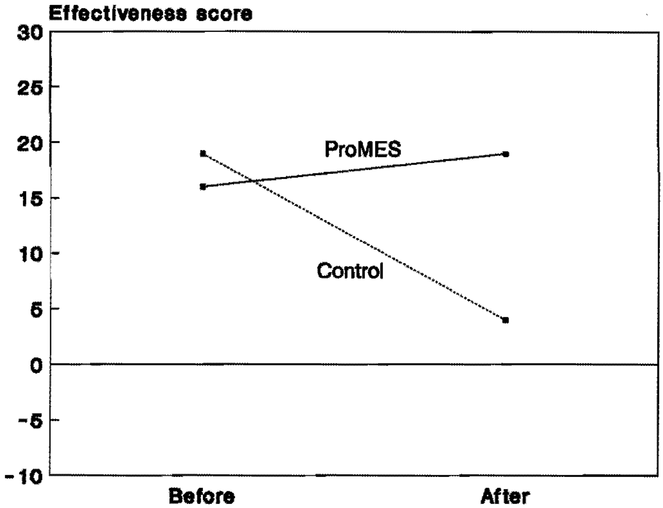


Figure 4.9. Individual indicators: summary of productivity changes.

It is interesting to compare the time series data from Figure 4.8 to the events that occurred. In the beginning of June, agreement was reached between design teams and management on the revised design, the tie between ProMES and the performance appraisal, and the institution of bilateral feedback meetings. The system was revised in July and August, and feedback resumed in September (feedback reports were then provided on June, July, and August concurrently). As mentioned before, no 'official' feedback meetings were conducted; feedback reports were sent to the technicians by mail, awaiting the design of a procedure for bilateral feedback meetings. This 'feedback by mail' continued through March 1992. The performance appraisal interviews, for which ProMES provided part of the input, were conducted in the second half of October 1991.

The first interesting finding pertains to the three-month period during which the system was being revised. Even though no ProMES feedback could be supplied, the technicians in the experimental regions increased their productivity compared to the control group technicians (32 versus 22 ProMES points). A possible explanation for this finding is an increase in the perceived importance of the feedback and the desirability of productivity improvement, due to the extensive discussion with management and the link between ProMES and the performance appraisal system. The absence of timely feedback may have been compensated for by feedback from previous months which may have directed attention toward areas (indicators, types of copiers) that could be improved. When looking at the four indicators individually, it is not surprising that the largest increase relative to the control group occurred with indicator 'Labor time per call'. Immediate improvements on this indicator may be obtained by 'simply' expending more effort (work faster), which is less complicated than trying to devise strategies to improve quality or reduce costs through 'working smarter'.

Secondly, it is interesting to note that the relative productivity increase of the experimental region compared to the control region is in fact a sharp decline in control group productivity compared to a productivity level in the experimental group which remains approximately the pre-revision level (this can be derived from the MANOVA and t-test in Table 4.16). When discussing the declining productivity with the design teams and management, the prevailing opinion was that the primary cause was a re-division of the regions in which the number of service regions was increased from 11 to 14, which took effect at the beginning of September 1991. This regrouping of technicians meant that a lot of technicians had to get used to new colleagues, a new supervisor, and less familiar clients and photocopiers. This may have caused the three-month productivity slump starting in September 1991.

To assess the impact of the re-division of regions on the experimental and control group, two measures were calculated: the percentage of technicians with more than 50% new colleagues, and the percentage of technicians with a new supervisor. These are shown in Table 4.17.

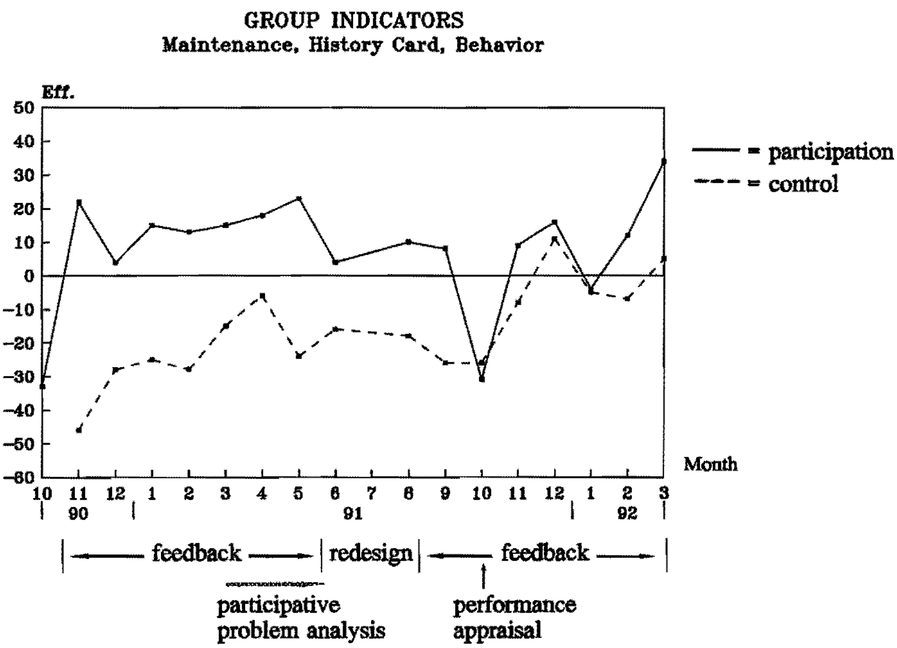
Table 4.17. Consequences of re-division of regions.

measure	ProMES	Enschede	Utrecht	control	overall
more than 50 percent new colleagues	32%	26%	38%	20%	22%
new supervisor	67%	26%	100%	40%	45%

The results show that the technicians from the ProMES condition were affected somewhat more by the re-division of regions than their colleagues in the control condition. So, the increased productivity in the ProMES condition was not the result of more favorable 'working-conditions' after the re-division of regions. On the contrary, the region that contributed most to the productivity increase in the ProMES condition (Enschede) was less affected by the re-division of regions than the region that contributed less (Utrecht). This suggests that the overall increase in the ProMES condition relative to the control condition might even have been a bit higher if no re-division of regions had taken place.

4.8.2. Group indicators

The group indicators are important in assuring that any productivity improvement on the individual indicators is obtained fairly, for a temporary improvement in cost could be obtained at the expense of long-term quality (as explained in Section 4.4.2). Figure 4.10 shows that the overall group effectiveness score of Enschede/Utrecht remained superior to that of the control condition (respectively 6 and −10). Whereas all of the three contributing indicators showed a significant positive difference during the pilot feedback period, in the post redesign feedback this was not always the case due to increased variance in the results of the experimental group (see Table 4.18).



Note: because of the limited amount of data collected in the holiday period (July and August), results of these months are combined and shown under August.

Figure 4.10. Group indicators: effects.

Table 4.18. Comparison of group effectiveness between ProMES and control regions.

	Enschede/Utrecht	Control regions	paired t-test (2-tailed)
Preventive maintenance	−6	−20	P=.048
History Card	−4	−3	P=.699, ns
Behavior	16	13	P=.482, ns
Overall group eff.	6	−10	P=.003

4.9. THE FINAL DESIGN

At this point, we will present a concise description the final design of the ProMES system. This description will serve both as a conclusion to the first phase of the design process described in this chapter, and as a starting point for the second phase of the design process, described in the next chapter.

Contingencies

A complete set of contingencies for type 4100 is shown in Figure 4.11 (next page).

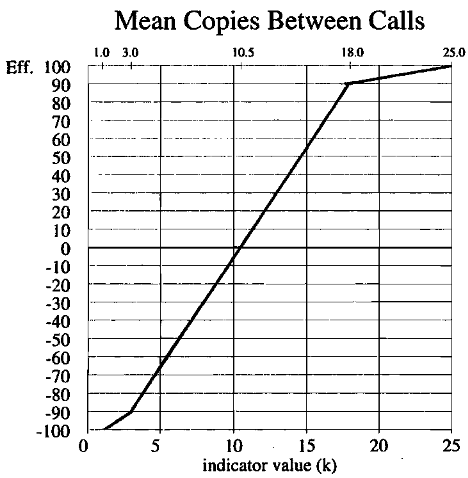
The individual feedback report

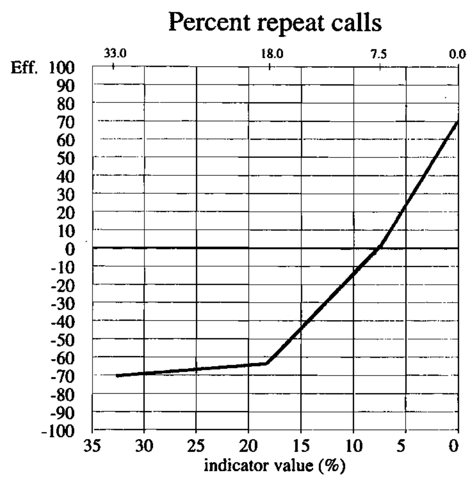
The individual feedback report consists of 3 pages: 1) a summary overview of individual effectiveness (Figure 4.12); 2) a detailed overview per type of copier of individual effectiveness (Figure 4.13); 3) a 'for your information' section, containing indicator values of the indicators which were deemed insufficiently valid. (No effectiveness scores are calculated; these results do not contribute to the technician's overall effectiveness score) (Figure 4.14).

```
                     PROMES INDIVIDUAL FEEDBACK REPORT

Region     : 1
Technician : M. Tenance
Period     : January 1993

           SUMMARY OVERVIEW OF INDIVIDUAL EFFECTIVENESS

                                   month    moving average
QUALITY (short term)
Mean Copies Between Calls           28           7         (-100,100)
Percentage repeat calls             22           1         ( -70, 70)
                                   -----       -----
                                    50           8
COST
Parts cost per call                 10-          5         ( -50 ,50)
Labor time per call                  9-          4-        ( -85, 85)
                                   -----       -----
                                    19-          1
                             ===============================
Overall individual effectiveness    31           9
```

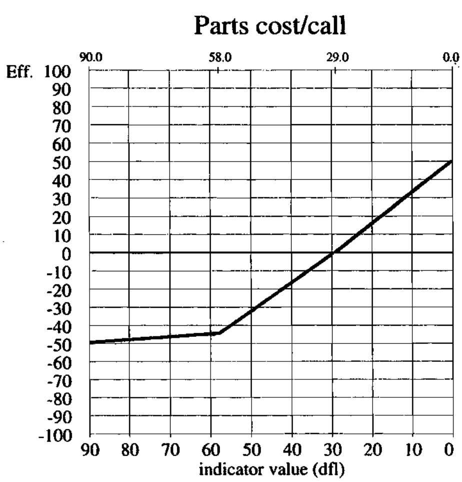
Figure 4.12. Individual feedback report page 1: summary overview of individual effectiveness.

Quality





Cost



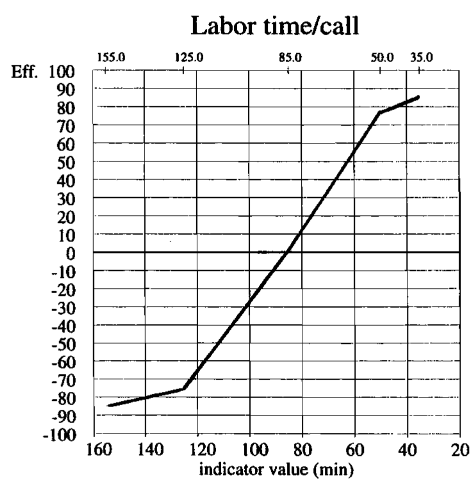


Figure 4.11: Contingency set type 4100.

DETAILED OVERVIEW OF INDIVIDUAL EFFECTIVENESS

Region : 1
Technician : M. Tenance
Period : January 1993

	calls		Month value	Month effect.	Moving Average value	Moving Average effect.		calls
QUALITY (short term)								
Mean Copies Between Calls								
Type P [4100]	28		13.1 k	25	11.9 k	20		135
Type Q	1				19.4 k	12		13
Type X	14		14.9 k	1-	9.4 k	43-		85
Type Z	11		50.5 k	73	42.7 k	34		89
				28		7		
Percentage repeat calls								
		%rep					#rep	
Type P [4100]	33	2	6.1 %	13	8.4 %	4-	13	155
Type Q	4				6.7 %	22	1	15
Type X	7	2	28.6 %	52-	18.5 %	17-	15	81
Type Z	17	0	0.0 k	70	10.0 %	23	9	90
				22		1		
COST								
Parts cost per call								
Type P [4100]	33		28.50.-	1	21.60.-	13		155
Type Q	4				30.10.-	7-		15
Type X	7		9.30.-	29	17.50,-	7		81
Type Z	17		147.50.-	46-	67.00.-	8-		90
				10-		5		
Labor time per call								
Type P [4100]	33		93 min	15-	86 min	9-		155
Type Q	4				105 min	46-		15
Type X	7		81 min	34	95 min	6		81
Type Z	17		125 min	15-	120 min	8-		90
				9-		4-		

Figure 4.13. Individual feedback report page 2: detailed overview of individual effectiveness.

INDIVIDUAL INDICATORS: FOR YOUR INFORMATION

Region : 1
Technician : M. Tenance
Period : January 1993

	month	moving average
ADMINISTRATION		
Percentage claims		50.9%
ATTENDANCE		
Percentage of capacity used	99.0%	100.8%
PRODUCT KNOWLEDGE[9]		
Perc. return calls knowledge	0.0%	1.0%

COST

Percentage return calls parts car stock

	#calls	#ret			#calls	#ret
Type P [4100]	33	1	3.0%	3.9%	155	6
Type Q	4			0.0%	15	2
Type X	7	1	14.3%	4.9%	81	4
Type Z	17	1	5.9%	8.9%	90	8

Labor time covered by ProMES: 96 hours 01 minutes[10]

Figure 4.14. Individual feedback report page 3: 'For your information'

[9] 'Product knowledge' (a technician should make sure his technical knowledge remains up-to-date) was rejected as a product in ProMES, since the results on the 'Quality' and 'Cost' indicators would be good measures of a technician's product knowledge. However, feedback on the percentage of return calls due to lack of knowledge was considered useful (e.g. for determining whether a technician would need extra training). Therefore, this measure was included 'unofficially' in the ProMES report.

[10] The hours of labor time measured in the ProMES report are an indication of the amount of work measured by ProMES. On average, technicians spend approximately 100 hours of labor time per month. Since not all types of photocopiers are included in the ProMES feedback, some of the technicians only receive feedback on a part of their work, so the hours of labor time will be less than 100. Likewise, when a technician is absent due to illness of vacation, he will receive a limited amount of feedback in the month in question, again resulting in a lower number of labor time hours. In practice, 25 hours of labor time was considered the minimum amount needed to have use for the feedback report. Feedback based on less than 25 hours is less useful because of an increased variance due to external factors (e.g. one repeat call could result in an effectiveness decrease of more than 100 points).

The region feedback report

The one page region feedback report (Figure 4.15) consisted of three sections: 1) feedback on the group indicators. measured by the supervisor (not included in the individual feedback report); 2) feedback on the individual indicators, which is the weighted average of the effectiveness scores of the region's individual technicians (the weight being the number of hours labor time included in ProMES); 3) feedback 'for your information' on the less than valid indicators.

PROMES REGION FEEDBACK REPORT

Region : 1
Period : January 1993

------- GROUP INDICATORS -------

	month		moving average		
QUALITY (long term)					
Maintenance procedure	87.3%	19-	89.8%	2-	(-70, 45)
ADMINISTRATION					
Accuracy of History Card	96.0%	3	90.8%	12-	(-45, 15)
REPRESENTATION					
Correctness of behavior	98.7%	26	95.0%	15	(-30, 30)
Overall effectiveness:		10		1	

------- INDIVIDUAL INDICATORS -------

	month	moving average	
QUALITY (short term)			
Mean Copies Between Calls	4-	9-	(-100,100)
Percentage repeat calls	16	10	(-70, 70)
	12	1	
COST			
Parts cost per call	1-	2	(-50, 50)
Labor time per call	11	16	(-85, 85)
	10	18	
Overall effectiveness:	22	19	

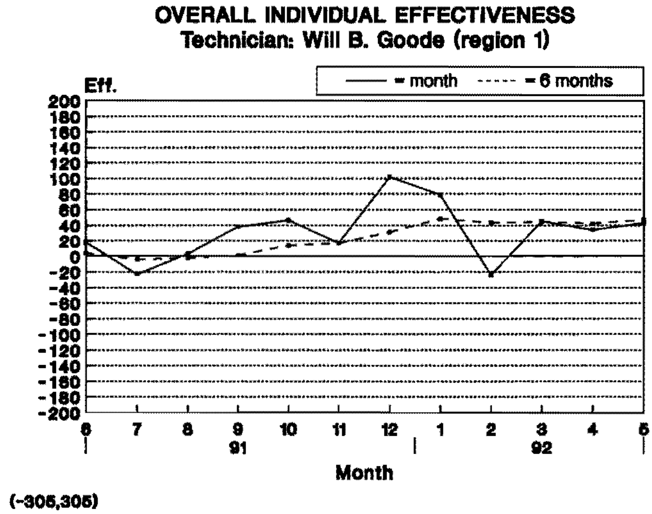
------- FOR YOUR INFORMATION -------

	month	moving average
ADMINISTRATION		
Completeness of claims	27.4%	15.3%
ATTENDANCE		
Percentage of capacity used	100.6%	99.7%
PRODUCT KNOWLEDGE		
Perc. return calls knowledge	0.3%	0.4%

Figure 4.15. Region feedback report.

Additional graphic feedback

The graphic feedback consisted of five graphs: one overall graph and one graph each for the for individual indicators, showing the monthly effectiveness and the six-month moving average of the previous 12 months up to and including the current month. See Figure 4.16 for an example.



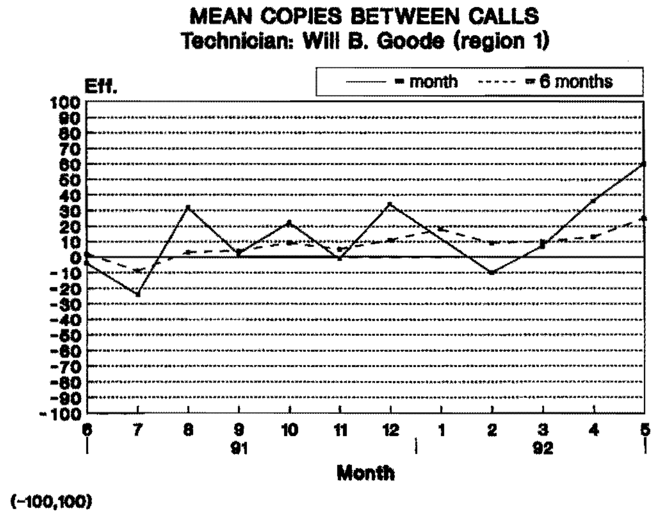


Figure 4.16. Additional graphic feedback.

4.10. CONCLUSIONS

We will conclude this chapter with some comments regarding the first design question ("How can an accepted control loop be designed in the Nashuatec setting with its specific characteristics?") from two perspectives. First, did the system contribute to the attainment of the company's objectives? Secondly, did the design process in this setting result in an acceptable design which confirmed to the design criteria for an accepted control loop?

4.10.1. Were the company's objectives met?

At the start of the ProMES program, management hoped the system would contribute to three objectives: improving the productivity of service technicians, supplying them with meaningful feedback, and (in the long run) presenting accurate quantitative input for their performance appraisal (see Section 3.2.3). Notwithstanding some serious problems during design and (pilot) implementation of the system, these objectives have been met for the technicians involved.

Positive effects on productivity could be established, first for the group indicators, and later on for the individual ones. Although the individual improvements that occurred after the system had been revised were not very large, the parties concerned were confident that a further increase would be obtained as soon as the bilateral feedback meetings–the last solution from the participative problem analysis–would be instituted. The second objective–providing the technicians with meaningful feedback–was also met. A positive result in this respect was the high opinion the management had of the measurement part of the ProMES system: they were convinced that the complexities of the job of the service technicians were accounted for extremely well by the system and that, after the revisions described before, the system was valid. They were particularly enthusiastic about the way a previously unsolvable problem, viz. the incomparability of performances of technicians on different types of photocopiers, was taken care of by introducing a separate set of contingencies for each type of copier. Not only the management, but all parties concerned were of the opinion that the measurement system offered by ProMES was a sound one, and far better than any means of feedback used in the past, thereby contributing to the attainment of the third objective, the use of ProMES data as a partial basis for the technicians' performance appraisal and distribution of the year-end bonus.

4.10.2. Did the design process result in an accepted control loop?

The setting in which the system was designed and implemented possessed a number of less desirable characteristics from the viewpoint of achieving a successful design and implementation of an accepted control loop. The complexity of the work flow and the existence of interdependencies between individual technicians combined with the possible future use of the system for performance appraisal purposes placed special demands on the measurement system and caused a high complexity of the system. The lack of vertical and horizontal communication combined with the top down management culture had created a situation of distrust and disrespect between regions and management, which seriously hindered the initial stages of the design process. The absence of visible management support of and commitment to the system during the pilot feedback period created a feeling of dejectedness within the design teams, that was alleviated only when management showed its willingness to confront all the problems generated by the design teams in this period. In the end, by creatively applying the ProMES guidelines, the majority of these problems–without exception referring to feedback criteria, goal criteria or context criteria that were not met–were overcome and a control loop resulted that met most criteria.

The main factor responsible for these results seems to be the participative approach advocated by ProMES and scrutinously followed in this case. Both the identification of the major causes for the lack of productivity improvement and the discovery of solutions for those causes resulted from that approach. Solutions that appeared to be decisive were those that made the system acceptable to the technicians. Of prime importance in this respect was the measurement system's perceived validity, caused both by the way data from different types of copiers were combined into meaningful scores and by the management's concession to remove apparently invalid indicators from the system. Another important factor was the link that was installed between performance and rewards on the request of the technicians, which increased the perceived importance of the feedback. The technicians experienced a lot of recognition from the part of the management by getting their problems accepted as real ones and by being taken seriously in the process of finding solutions. Taken together, this resulted in a system trusted by those who had to work with it.

4.10.3. Continuation of the design process

An important precondition that the otherwise acceptable ProMES design did not yet fulfil was its suitability for setting accepted and challenging goals for all technicians through 'individualized' feedback meetings (the sixth cause in Section 4.7.2). This was the result of the procedure that had to be used in establishing the contingencies. Feedback meetings between individual technicians and their supervisors, attuned to the specific abilities and circumstances of individual technicians were considered a prerequisite for effective goal setting to take place. To provide the supervisors with the knowledge and skills necessary to conduct such feedback meetings, a feedback training–especially designed for the Nashuatec context–was executed. The design of the feedback training and the individual feedback meetings is described in Chapter 5.

The main subject of the next chapter, however, is the transportation of the ProMES system to six regions that did not take part in its design. Since the participative approach taken in the design process was considered essential to the system's eventual success, it remained to be seen whether a control loop which was successful in the design region will also improve the productivity of regions that did not have the opportunity to participate.

Chapter 5
Non-participative implementation of a ProMES system in six regions

5.1. INTRODUCTION

The previous chapter provided a positive answer to the first research question concerning the possibility of successfully designing and implementing a ProMES control loop in the Nashuatec setting. The implementation of the participatively designed system in the two design regions resulted in significant productivity improvement. This positive result paved the way for the process of finding an answer to the second main research question of this research project:

"In this setting, can the design be successfully implemented in other–comparable–groups, without going through the participative design process?"[1]

After the participative design and redesign processes had produced a ProMES system which functioned satisfactorily, this research question was answered through a non-participative implementation of the system (designed participatively by the Enschede and Utrecht regions) in six of the remaining twelve regions. As shown in Section 3.5.2, this implementation corresponds with an 'own decision with explanation' decision-making strategy. The transportation[2] process will be set out in detail in Section 5.2. Its effects on productivity and cost reduction will be evaluated in Section 5.3. and 5.4 respectively. In Section 5.5, the productivity effects of the participative approach and the transportation approach are compared. Finally, conclusions will be drawn in Section 5.6.

5.2. DESIGN OF THE TRANSPORTATION PROCESS

5.2.1. Introduction

The non-participative implementation of ProMES in six regions comprised four main elements (see Figure 5.1). First, a standardized introductory meeting provided the technicians with a basic introduction into the ProMES system which they were to use (Section 5.2.3). Secondly, follow-up to this introductory meeting was provided by means of bilateral feedback meetings between the region supervisor and his individual technicians (Section 5.2.5). As mentioned in Chapter 4, an individualized feedback and goal setting procedure was considered a prerequisite for optimum use of the ProMES

---

[1] The relevance of this research question has been explained in Chapter 3.

[2] The non-participative implementation process will be labelled *transportation* throughout this chapter (after Pritchard et al., 1989).

control loop. In addition, the bilateral feedback meetings were to serve as an opportunity for the technicians to get a more detailed understanding of the system and as a means for arriving at some degree of acceptance of the system. In order to provide the supervisors with the knowledge and skills necessary to conduct effective feedback sessions, a feedback and goal setting training program, geared to the specific characteristics of the ProMES system at Nashuatec, was designed (described in detail in Section 5.2.4). Ten months after the introductory meeting, the implementation process was concluded with an evaluation meeting (see Section 5.2.6).

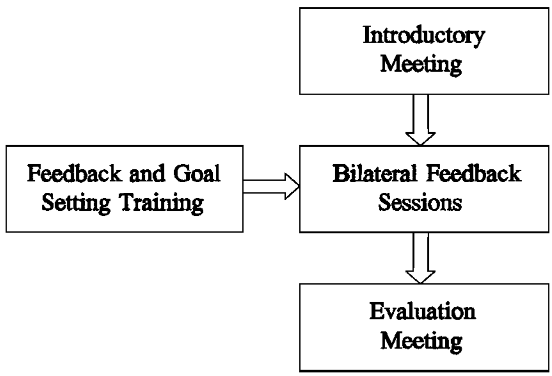


Figure 5.1. Design of the transportation process.

5.2.2. Choice of regions

Criteria

Adequate assessment of the effects of transporting the ProMES system is feasible only if some of the regions remain control groups; these are regions in which ProMES is not implemented. The effects of transportation are assessed by comparing (among others) the ProMES effectiveness scores of the transportation regions with those of the control regions. After consultation with management and supervisors, the decision was made to transport the ProMES system to six regions, and to use the remaining six regions as control groups for a one year period.

Three criteria were employed in establishing which of the 12 remaining regions should be allocated to the transportation condition and which to the new control condition.

The first criterion stems from (quasi-) experimental research guidelines. When using a design with control groups, it is imperative that the experimental groups (in this case: transportation groups) and control groups are matched, i.e. 'made equal' on variables that may influence the outcome of the experiment (e.g. Janssens, 1988). The first matching variable used was the average ProMES effectiveness score of the 12 regions in a 10 month baseline period before the start of the transportation process. To insure that differences in productivity changes between the transportation and control condition would not be (partially) caused by differences in room for improvement at the start of the transportation process, the average overall ProMES effectiveness score should be about equal for both conditions. The second matching variable concerned the regions' geographical position in the country (urban versus rural), since there were some indications that the distinction between an urban and a rural environment represented differences in attitudes and work situations (imagine for example a technician's typical workday in the center of Amsterdam with one in the more rural environment of some locations in the East of the Netherlands). During the design phase of the project, the importance of this criterion was underlined by the rather positive attitude in the rural region of 'Enschede' and the rather negative attitude encountered in the urban region of 'Utrecht'.

The second criterion dealt with the willingness of the supervisors to spend extra time on the ProMES program in the coming period. Because the transportation of ProMES would entail a lot of effort on the part of the supervisors (e.g. participation in the training program, bilateral feedback sessions), they were given the opportunity to decide for themselves whether they would want to invest this effort the coming year or the year after that. Therefore, only those regions whose supervisors either preferred direct involvement in the ProMES program or expressed no preference would be eligible for the transportation condition.

The third criterion was a practical one: because of a re-division of regions, technicians of the Utrecht region who had originally participated in the development of ProMES, had been relocated to two newly formed regions. Since it would not be justified to supply only part of a region with ProMES data, it was decided to introduce ProMES to the other technicians of these regions in the framework of the transportation process[3], so that the regions would remain homogeneous in this respect.

Procedure

During a meeting at which all supervisors and the two field service managers were present, the final version of the ProMES system (depicted in Section 4.9) was presented and discussed. The supervisors were then asked whether they preferred direct involvement in the transportation of the system or preferred postponing it until next year. In response, one supervisor preferred to wait, two supervisors had no preference and the rest would like to start directly.

---

[3] This group of technicians (made up of technicians from two different regions) is called region 'Hoevelaken' in this chapter. This 'pseudo-region' also included a few technicians who were transferred to the region of 'Enschede'. On average, the attitude of these technicians was similar to that of technicians who did not have any colleagues in the region that had participated in the design of the ProMES system. Therefore, the technicians from 'Hoevelaken' were included in further analyses.

Through a combination of all three criteria (matching of conditions on baseline effectiveness and urban-rural distinction, homogeneous regions, and supervisors' preferences), the allocation was done by the researcher, after consultation with the service manager (see Table 5.1).

Table 5.1. Allocation of regions to transportation and control conditions.

TRANSPORTATION CONDITION	Criterion 1		Criterion 2
	Baseline effectiveness	Geographical Position	Preference of supervisor*
Region 4: Arnhem	25	Rural	+
Region 6: Eindhoven	19	Rural	+
Region 3/12: Hoevelaken (crit. 3)	−1	Urban	+
Region 11: Den Haag	−1	Urban	+
Region 7: Maastricht	−9	Rural	+/−
Region 8: Haarlem	−9	Rural	+
	Mean eff. = 3	2 urban, 4 rural	5+, 1+/−, 0−

CONTROL CONDITION	Criterion 1		Criterion 2
	Baseline effectiveness	Geographical Position	Preference of supervisor*
Region 1: Groningen	19	Rural	+
Region 5: Den Bosch	18	Rural	+
Region 9: Amsterdam	18	Urban	+
Region 14: Breda	0	Rural	+/−
Region 13: Rotterdam	−2	Urban	−
Region 10: Leiden	−23	Urban	+
	Mean eff. = 5	3 urban, 3 rural	4+, 1+/−, 1−

* '+' = direct involvement; '+/−' = no preference; − = postponed involvement

Table 5.1 shows the allocation of the 12 regions to the transportation and control condition. The transportation and control condition were successfully matched according to baseline effectiveness (in the post-redesign period), the mean effectiveness value being 3 in the transportation condition and 5 in the control condition[4]. This two-point difference is very small when considering the overall range of effectiveness values (from 25 (Arnhem) to −23 (Leiden)). The second matching requirement was fulfilled: both conditions contain a similar number of urban and rural regions.

[4] The actual period used for matching the transportation and control condition was May through October 1991 (preparation for the transportation process started in November 1991). Table 5.1 shows that the conditions were actually matched at the start of the transportation process.

The second criterion hardly limited the freedom of allocation, since only one supervisor from an urban region preferred postponement of his region's involvement in ProMES (which was accordingly allocated to the control condition). In accordance with the third criterion, homogeneous regions were formed by including region Hoevelaken in the transportation condition.

Two months before the start of the transportation, briefly worded information was sent to technicians of all twelve regions concerned, regarding the objectives, design, contents, and application of the ProMES system at Nashuatec. This was accompanied by a timetable for the implementation in the 12 regions.

5.2.3. Introductory meeting

The first step in the transportation process consisted of a three hour introductory meeting in each of the transportation regions, conducted during six evenings within a one month period. The purpose of the introductory meeting was to initiate the process of convincing the technicians that attention for their productivity is a legitimate concern, and that ProMES may be a useful tool for monitoring their productivity. At the same time, the meeting should result in the technicians having a basic understanding of the system when they started the bilateral talks with their supervisor. Present at the meetings were the technicians and supervisor of the respective regions, two management representatives (the field service managers responsible for the region and the service specialist from the personnel department) and the two facilitators.

The introductory meeting contained five main elements: a management statement, a general introduction into performance management theory and ProMES, an introduction into the Nashuatec system, a round of questions and answers, and the distribution of materials to be used. These are described briefly.

Management statement

The meeting started with a 10 minute opening statement from the field service manager. In this statement, he stressed the importance of attention to productivity improvement of the service technicians for surviving in the highly competitive photocopying market. The detailed maintenance procedures and product training programs were mentioned as examples of management's concern for the technicians' productivity. Attempts by management to provide the technicians with accurate performance feedback were mentioned, as were the continual failures to do so effectively due to a number of shortcomings. These shortcomings included incompleteness or oversimplification of measurement, insufficient controllability of the performance-indicators, and inaccuracy of the data fed back. Next, ProMES was presented as a system which incorporated solutions to most of the problems encountered in the past and which therefore provided more accurate and useful feedback than any system previously used. As important contributing factors to this success, the high degree of participation by the technicians and supervisors from Enschede and Utrecht was mentioned as were system features (e.g. contingencies) to be illustrated later on.

The technicians present were assured that ProMES was not intended as a monitoring system for higher management; on the contrary, the technicians should regard the system as their own tool for monitoring their own performance and as a basis for constructive talks with their supervisors about their job.

The field service manager concluded his introductory statement by expressing his support for the system and his confidence that using ProMES would enable the technicians to contribute to the company's success even more effectively.

General introduction of performance management theory and ProMES

In a 45 minute presentation, the concepts of feedback and goals, essential to self-regulation, were explained in accessible terms, as were the requirements for an effective feedback and goal setting procedure. The necessity of a good productivity measurement system as a basis for conducting a feedback and goal setting program was highlighted.

Next, the general ProMES design procedure was explained by means of a simplified example (adapted from the ProMES system used by a corrugated fibreboard production team (Van Berkel, 1990)). Special attention was drawn to the positive measurement properties of the system, such as a single overall index of productivity, direct comparison of different indicators, and representation of the relative importance of indicators in the system.

Introduction of the Nashuatec system

A detailed 90 minute description of the ProMES system at Nashuatec was given by the researcher. The main topics of this description were:
* reasons for starting the project;
* the design procedure in Enschede and Utrecht;
* the products and indicators developed (with special attention to the level of measurement [individual versus group level], the type dependency issue, and the indicators rejected due to validity problems);
* the contingencies (with special attention to the way the contingencies solve the type dependency, the use of both cost reduction and customer satisfaction criteria to determine the relative importance of the indicators, and the use of historical data to establish the range of indicator values);
* the feedback report and the feedback graphs;
* effects of the implementation of ProMES in regions Enschede and Utrecht;
* use of the system (bilateral feedback meetings);
* performance appraisal (procedure used in Enschede and Utrecht).

Questions and answers

The technicians were given the opportunity to ask informative questions during and after the above presentation. The questions mainly focused on measurement issues (e.g.

"How did you make sure that all supervisors judge our performance on the region indicators in the same way?") and issues concerning 'fair' use of the system (e.g. "Will the zero point be changed if our productivity improves?"; "What information will higher management receive on our individual productivity?"). In all regions, the system was provisionally given the benefit of the doubt: "The system looks good, but let's try it to find out whether it is really as accurate and useful as you say".

It was decided to postpone the issue of possible future use of ProMES for performance appraisal purposes to the evaluation meeting. This would give the technicians the opportunity to form an informed opinion on whether the system provided performance data that would increase performance appraisal accuracy.

ProMES manual and feedback reports

At the end of the meeting, a ProMES manual (Kleingeld & Van Tuijl, 1992) was distributed. This manual, written and compiled especially for the technicians and supervisors at Nashuatec, contained 5 sections: 1) the ProMES method: an overview by example, 2) the ProMES system at Nashuatec: a reference guide, 3) the ProMES feedback report: an overview of calculations, 4) the region indicators: checklists used by supervisors, and 5) the current set of contingencies. The manual included an introductory section, in which management repeated a number of issues they brought up during the introductory meeting. The purpose of this manual was twofold: providing technicians (and supervisors) with an immediate reference for most of their questions and serving as a file for all information the technicians would receive regarding ProMES (monthly feedback reports, new sets of contingencies, memorandums, etc.). Technicians were asked to have the manual with them in their company car at all times.

Finally, all technicians received copies of their individual feedback report and the region feedback report of the last three months (previously only available to the facilitator) in order to get a first impression of the type of feedback they would be receiving in the future and the level of productivity at which they had been performing.

5.2.4. The feedback and goal setting training program

Background

One of the main perceived causes of the lack of more substantial productivity improvement in the regions that participated in the design of ProMES was the ineffectiveness of the group feedback meetings (Section 4.7.2). Because a large part of the system focused primarily on the productivity of individual technicians, it was proposed to discuss the feedback reports in bilateral meetings of an individual technician and his supervisor. In such meetings, the knowledge and skills and circumstances of each individual technician could be taken into account more effectively (e.g. when setting specific, challenging goals). These meetings had not been implemented, because it was felt that one should first ascertain that the supervisors possessed the knowledge and skills to conduct these meetings effectively.

There were two basic considerations for designing and conducting a training program. First, when providing feedback, certain leadership styles are considered more effective than others in stimulating employees to accept discrepancies between feedback and (implicit or explicit) goals, and engage in activities to reduce these discrepancies, resulting in improved performance (e.g. Latham & Saari, 1979a, 1979b; Latham & Wexley, 1981). In this way, the training should help the supervisors improve their leadership skills in the area of performance management. Observations during feedback meetings and on other occasions indicated that the leadership style of at least part of the supervisors differed from the 'problem-solving' leadership style considered suitable within a ProMES program.

The second consideration was a practical one: when transporting the ProMES system to regions that had not participated in its development, problems resulting from a lack of understanding of the system and/or resistance to the objectives or approach of the program, could be expected to occur. Supervisors should be able to deal with these issues effectively before starting the feedback and goal setting procedure. The training program should therefore concentrate on dealing with resistance and providing explanation, in addition to discussing the feedback reports and setting goals.

Behavior modeling

In the training program, elements from the 'behavior modeling' approach to improving supervisory skills (e.g. Decker & Nathan, 1985; Goldstein & Sorcher, 1974; Latham & Saari, 1979b) were incorporated. This approach, based on Bandura's social learning theory (Bandura, 1977), typically consists of the following steps (e.g. Latham & Saari, 1979b):

1) introduction of the topic by the trainer;
2) presentation of a film that depicts a supervisor model effectively handling a situation by following a set of 3 to 6 learning points that are shown immediately before and after the model is presented;
3) group discussion of the effectiveness of the model in demonstrating the desired behaviors;
4) practice in role playing the desired behaviors in front of the entire class;
5) feedback from the class on the effectiveness of each trainee in demonstrating the desired behaviors.

Taken together, these steps include the attentional, retention, motor reproduction, and motivational processes that are considered components of effective modeling (Bandura, 1977).

The main difference between the above list of 'behavior modeling' activities and the procedure followed in the Nashuatec training program is the lack of a visual model to demonstrate the desired behaviors. Due to the specialist nature of the key learning points to be used in the Nashuatec training program, only a custom-made film would have been a useful modeling tool. Unfortunately, making such a film would have been too costly and time-consuming. So, the training program did not represent a complete 'behavior modeling process'; apart from this, the guidelines were followed closely.

Contents of the training program

The supervisors of the two regions involved in the participative design of the system and those of the six transportation regions took part in the training program. The official objective of the program was: "Conveying to the supervisors the necessary knowledge and skills to conduct a ProMES feedback session in such a way that technicians will use ProMES as a self-regulatory system" (Van Tuijl & Kleingeld, 1992). The contents of the program, described below, are summarized in Table 5.2.

A. Day One

Presentation of feedback and goal setting principles. Based on the accepted control loop model (elucidated in Section 2.4.1), the ProMES measurement system was presented as the basis for effective feedback and goal setting (and reinforcement). The requirements for effective feedback and goals were discussed, with special attention to the contribution of the bilateral feedback meetings to conforming to the requirements that would not be met otherwise (e.g. setting difficult but attainable goals for each individual technician).

Technicalities. Due to the complexity of the ProMES system, some characteristics may confuse users if they are not explained properly. Since the supervisors would be the ones confronted with 'technical' questions during the feedback sessions, they should be prepared properly. Therefore, the main technicalities of the system were discussed.

Presentation of Key Learning points. Five sets of Key Learning Points[5] (i.e. recommended guidelines) were developed to aid the supervisors in conducting effective ProMES feedback meetings. The Key Learning Points represent a participative problem-solving style of leadership (Latham & Saari, 1979a; Latham & Wexley, 1981).

Before an effective feedback meeting can be conducted, two preconditions need to be fulfilled. First, agreement should be reached between technician and supervisor on the objective and contents of the procedure. Technicians, having been introduced to the ProMES system, will differ in the extent to which they accept the main objective of the ProMES approach–productivity improvement through increased motivation and better work strategies of personnel–and the extent to which they consider bilateral talks with the region supervisor a useful aid for attaining this objective. Secondly, the extent to which technicians understand the system may vary considerably (which may in turn cause resistance, e.g. because of perceived inaccuracy of the system). Consequently, the supervisor should find out to what extent each individual technician accepts and understands the system. In case either acceptance or understanding is low, the supervisor should first work out these issues with the technician. The accompanying Key

---

5 A number of sources were used in drawing up the Key Learning Points. These sources include: Goldstein & Sorcher (1974), Latham & Saari (1979b), Latham & Wexley (1981), Pritchard (1990), and Video Arts (1988).

Learning Points are listed in Table 5.3. Only if the technician accepts and understands the system to a reasonable degree, should the supervisor commence the next phase: discussing the feedback reports.

Table 5.2. Overview of the feedback and goal setting training program.

FEEDBACK AND GOAL SETTING: A TRAINING PROGRAM	
A. Day One	1. Presentation of feedback and goal setting principles. 2. Technicalities: discussion of characteristics of the company's ProMES system, relevant for its application. 3. Presentation and discussion of Key Learning Points - Phase 1: Presenting objectives and procedure/ dealing with resistance; - Phase 2: Feedback; - Phase 3: Goal setting; - Discussion Until Consensus; - Maintaining a constructive atmosphere. 4. Role-playing sessions (phase 1 and 2). 5. Constructive feedback on the effectiveness of the trainee in demonstrating the desired behaviors. (5 hours, two groups of 4 supervisors, 2 trainers)
B. A month of practice	On-the-job use of skills learned: * All trainees conduct at least five real-life feedback meetings during a one month period. * Each feedback meeting is evaluated using an evaluation form.
C. Day Two	1. Evaluation of on the job experiences with the feedback meetings. 2. Role playing exercises: - trying out approaches for difficulties encountered; - goal setting. (5 hours, two groups of 4 supervisors, 2 trainers)

The second phase of a bilateral feedback meeting entails the discussion of feedback reports. If the feedback reports are discussed for the first time, the supervisor should first explore whether the technician has problems understanding his feedback report or has doubts about the accuracy of the feedback data. These issues should be resolved before continuing this phase. The major topics in the discussion of the feedback reports are: examining in which areas (products, indicators, types of copiers) the technician performs well or less well (is satisfied or less satisfied), finding causes of high or low productivity (satisfaction or dissatisfaction), generating solutions, reaching agreement on implementing those solutions that lead to improved productivity, and monitoring progress. The accompanying Key Learning Points are presented in Table 5.4.

Table 5.3. Key Learning Points for 'Explanation of objectives and procedure' and 'Dealing with resistance'.

PHASE 1: EXPLANATION OF OBJECTIVES AND PROCEDURE & DEALING WITH RESISTANCE

1. Explain:
   a. the purpose of the talk;
   b. the necessity of the program;
   c. the objectives of the program;
   d. the contents of the program;
   e. prior conditions to the program.
2. Ask the technician's opinion about (1).
3. Listen openly to the technicians concerns about (1).
4. Verify whether you understand what the technician means.
5. Summarize the technician's remarks in two categories: agrees/doesn't agree.
6. Ask the technician to think about solutions (self/others).
7. Assess the results of (1)-(6).
8. Depending on (7): continue with phase 2 or plan a follow-up meeting.

The final phase of the feedback meeting consists of setting specific, challenging goals. The ideal starting point for this process would be the following: 1) the technician understands, recognizes and accepts the feedback reports, 2) he is not entirely satisfied with his productivity, 3) he has some ideas about how to improve his productivity, and 4) he has a high subjective probability of succeeding. Guidelines for this process of goal setting are depicted in Table 5.5.

The second and third phase are part of an iterative process. The goals set and strategies agreed upon during a meeting are used as input for the next meeting, during which the feedback from the intervening period is analyzed.

Table 5.4. Key Learning Points for 'Feedback'.

PHASE 2: FEEDBACK

1. Explain the purpose of this phase.
2. Ask the technician:
   a. whether he understands his report;
   b. to what extent he recognizes the report as a reflection of his performance.
3. Ask the technician to name causes of high/low performance (or: satisfaction/ dissatisfaction).
4. Ask the technician:
   a. to think about solutions to causes of low productivity/dissatisfaction;
   b. what he can contribute to these solutions;
   c. what others in the organization can contribute to these solutions.
5. Arrange the solutions according to attainability and expected degree of productivity improvement.
6. Jointly choose the best solution and means to implement it.
7. Agree upon a follow-up meeting to monitor effects.

Table 5.5. Key Learning Points for 'Goal Setting'.

PHASE 3: GOAL SETTING

1. Explain the purpose of this phase.
2. Ask the technician to describe his range of effectiveness scores of the past period.
3. Ask the technician to think about causes of fluctuations within this range.
4. Make a list of possible causes for low and high productivity.
5. Ask the technician for solutions for causes of low productivity.
6. Assess the attainability and expected effects of these solutions.
7. Choose the best solutions.
8. Ask the technician to establish *his own* maximum, zero point and minimum.
9. Ask the technician which effectiveness level would be attainable if the proposed solutions were implemented.
10. Make the joint decision that this effectiveness level will be the goal for the next period.

In addition to the Key Learning Points for the three main phases, general guidelines were drawn up for 'Discussion until consensus' and for 'Maintaining a constructive atmosphere' (see Tables 5.6 and 5.7).

Table 5.6. Key Learning Points for 'Discussion until consensus'

DISCUSSION UNTIL CONSENSUS

1. State your problem clearly.
2. Verify whether the technician understands the problem and is willing to cooperate in finding a solution.
3. Collect as many alternative causes as possible; postpone evaluation of these causes.
4. Evaluate all causes on the basis of criteria/argumentation.
5. Jointly choose the most important cause(s).
6. Ask the technician to think about solutions; postpone evaluation of these solutions.
7. Evaluate all solutions on the basis of criteria/argumentation.
8. Jointly choose the best solution(s).
9. Decide on the implementation (who, what, when) and set a follow-up-date.

Table 5.7. Key Learning Points for 'Maintaining a constructive atmosphere'

MAINTAINING A CONSTRUCTIVE ATMOSPHERE

1. Focus on issues, not on persons
2. Stress the positive
3. "Us against the figures"
4. Ask the benefit of the doubt
5. Propose a period of experimentation
6. Postpone the decision: "sleep on it"
7. Try a compromise: "give and take"
8. Ask advice of an impartial expert
9. Tell the truth: "no surprises"

Discussion of key learning points. The main difference between the procedure followed at Nashuatec and the 'ideal' behavioral modeling procedure was the absence of a visual model of effective supervisor behavior in the special setting of bilateral feedback meetings in the Nashuatec setting. So, this phase had to be limited to discussion on the advantages and disadvantages of these guidelines.

Role playing sessions. Next, several role-playing sessions were conducted, in which the desired behaviors were practiced. To make these sessions as realistic as possible, a large number of detailed scripts were developed, consisting of 1) a role description of a technician, including information on sources of resistance to the program and on the degree of understanding of the system, 2) a role description of a supervisor, containing brief information on the technician, and 3) feedback reports and graphs covering a six month period. Each time, one supervisor assumed the role of a technician based on this realistic script. Role playing concentrated on the first and the second phase of a bilateral feedback meeting. Using the Key Learning Points as guidelines the supervisor tried to deal with the situation as effectively as possible. Role playing sessions took between 15 and 25 minutes each.

Constructive feedback. After each role-playing session, the rest of the group provided constructive feedback on the effectiveness of the supervisor in demonstrating the key leaning points to deal with the situation that occurred. Alternative approaches to those taken in the role playing exercise were discussed and used as input for subsequent role playing.

B. On-the-job use of skills learned.

The trainees were asked to conduct at least five 'real life' feedback meetings during the next month, using the Key Learning Points as guidelines. To aid them in structuring their experiences, an evaluation form was provided, which consisted of a chronological overview of a bilateral feedback meeting divided into its main phases. For each phase a number of questions was asked regarding issues such as the technician's opinion of the approach taken, the extent to which the key learning points were followed, reasons for any deviations, use of available material, agreements reached, etc.. The on-the-job experiences were the basis for day two of the training program.

C. Day Two

Evaluation of on-the-job experiences with the feedback meetings. Based on the trainees responses on the evaluation form, the real-life meetings were discussed. In general, the trainees considered the Key Learning Points useful guidelines for conducting a bilateral feedback meeting.

A number of issues related to ProMES and its use in regular bilateral meetings came up during the second day of training. These will be discussed in Section 5.2.6 as part of the evaluation of the transportation intervention by the supervisors.

Role playing exercises. Since the first day role playing concentrated on the first and second phase of a ProMES feedback meeting, the majority of exercises during the second day concentrated on the (combined) second and third phase: discussing the feedback reports and setting goals. Roles were selected that contained problem characteristics similar to those encountered by the trainees in their on the job meetings. Since the discussion on issues related to the on-the-job experiences of the supervisors had taken up a lot of time, role playing was limited to only a few exercises.

5.2.5. Bilateral feedback sessions

The supervisors were allowed complete discretion in determining the frequency, duration and location of the sessions. It was, however, suggested that one meeting per three months for each technician might be a suitable frequency, because it may take a few months for effects of agreements implemented by the technicians to become visible in the ProMES reports. The supervisors were responsible for compiling and distributing the graphical feedback; the frequency of providing this feedback was up to each individual supervisor.

Management agreed to discontinue the annual 'test-appraisal' session (a kind of performance counseling interview on the basis of the performance appraisal form), because its problem-solving function could be performed more effectively in the context of a ProMES feedback meeting.

A few months before the start of the bilateral feedback sessions, the management had decided to work towards obtaining the ISO-9001 certificate for the service department, which was obtained in July 1992. Formalizing and describing the procedures within the service department required a lot of effort from the supervisors. The project also involved a number of extra meetings in the service regions. These activities were given a high priority and, therefore, the bilateral feedback meetings were given less attention than might otherwise have been the case.

5.2.6. Evaluation meeting

Procedure

Approximately ten months after the introductory meeting, a meeting was organized in all transportation regions with the purpose of discussing the technicians' experiences with the system until that date.[6]

The first 20 minutes of the meeting were used for filling in a questionnaire on the bilateral feedback meetings (to be discussed in Chapter 6).

After that, one of the field service managers made an opening statement which to a large extent covered the same ground as the one made during the introductory meeting. Among other things, he stressed that ProMES constituted a much more (albeit not 100 percent) valid measurement system than those used in the past. The usefulness of ProMES for technicians as a means for regulating their own performance was stressed too as was the role ProMES played as a common reference for technicians when talking about their work.

---

[6] A few months prior to this evaluation meeting, two regions in the transportation condition ('Arnhem' and 'Eindhoven') were visited by the researcher to answer a number of mainly measurement-related questions (at the request of the supervisor). Also, the researcher visited the regions that had participated in the design of the system to answer technical questions that had come up during the past six months.

Next, the technicians were asked to propose topics they would like to discuss. The first category should touch upon ProMES as a system to regulate the performance of service technicians (experiences with the system, strong and weak points of the system, and desirable alterations or additions). The contribution of ProMES to the performance appraisal of individual technicians (under which conditions would a direct link between ProMES and the annual bonus be acceptable?) was suggested as a second topic. These items raised by the technicians were divided into the two categories and prioritized. First, the measurement issues were discussed, then the performance appraisal issues.

The facilitator concluded the meeting with a short presentation of the results of a questionnaire regarding the use of feedback and goal setting in the field service department in general and in the region concerned (to be discussed in Section 6.2).

An interesting observation concerns the atmosphere during the evaluation meetings. There were large differences between the regions in this respect. In some regions, the technicians present were friendly, generally positive toward the system, and willing to accept compromises regarding imperfections of the system. In the majority of regions, on the other hand, technicians were distrustful of management and facilitator, very critical toward the system and unwilling to accept any imperfection of the system. An example of the latter category was the unwillingness of technicians of one region to supply the facilitator with a list of questions beforehand, because "he should not be allowed to prepare himself".

Main topics discussed

The following issues were considered important in most or all of the regions. They will be discussed briefly:

1. Influence of copy volume on MCBC
2. Limited controllability of some indicators
3. Perceived overestimation of the cost-product
4. Life cycle of photocopiers
5. Zero Point
6. Lack of information on causes and solutions
7. Long term effects/animosity between colleagues
8. Use of ProMES in performance appraisal

1. Influence of copy volume on MCBC. An issue that came up in the design stage of the project was the dependence of the indicator Mean Copies Between Calls on the amount of copies made per time period (labelled 'copy volume'). It is known that there is a non-linear positive relation between the number of copies a customer makes on a copier per unit of time and the MCBC result that can be obtained. Thus, in case of a high copy volume, the MCBC-score would be an overestimate of the quality-effectiveness of a technician; on the other hand, if the copy volume is low, the MCBC-score would be an underestimate. During the design of the system and its implementation in the two regions that participated in the design process, there were no large structural differences between regions and customers; if randomly assigned to customers (which is standard procedure in the planning department), technicians would encounter approximately the same copy volume. Since the relationship between the copy volume and MCBC is non-linear and not the same for each type of copier, taking it into account would increase the complexity of the indicator used. Both had been reasons for not taking into account differences in copy volume.

Since the implementation of ProMES in the spring of 1992, a number of external developments necessitated reconsideration of prior assumptions. These developments include: (1) the company's sales department began selling revised older copiers for locations representing a lower copy volume than these machines encountered in the past, (2) a new line of high-quality types of copiers 'drew away' copy volume from older types also used by the customer, (3) due to the economic recession in some sectors, some customers kept a tight hand over the amount of copies made.

Until a definitive solution for the copy volume-problem could be implemented, a temporary measure was taken, which involved a bi-annual calculation of the estimated change of MCBC due to change in copy volume and an according adjustment of the contingencies (in most cases a downward adjustment). This is an acceptable exception to the general rule that contingencies should not be changed following changes in effectiveness, since the change mentioned here is caused by external factors that are structural and uncontrollable and not by changes in effort or work strategies on the part of the technicians.

This periodical adjustment of the contingencies solved part of the problem. The MCBC-effectiveness scores obtained on types of copiers to which the decrease of copy volumes applies to a different degree remain comparable. However, technicians who frequently visit customers with a relatively low copy volume remain at a slight disadvantage. Latter problem can only be solved by directly relating the MCBC-score for each call to the copy volume of the machine in question. A less complex alternative would be adding an indicator such as 'Time between calls', which is negatively related to the copy volume.

2. *Limited controllability of some indicators.* An issue that came up often was the perceived limited controllability of indicators. Technicians argued that scores on most indicators were influenced by factors outside the control of individual technicians, such as customers, quality of spare parts, and colleagues. In his response, the facilitator divided these factors into two categories: factors that will affect all technicians to approximately the same degree and factors that will not.

The first group includes factors such as the location of the machine or machine malfunctions caused by the customer, the comprehensiveness of the maintenance procedure that has to be carried out (which influences spare part replacement and labor time), and the quality of the work done by a technician's colleagues who carried out the previous repairs and maintenance on a machine. Although these factors may be irritating and may cause short-term differences in circumstances between technicians, the fact that technicians encounter these factors randomly will limit the long-term differences. Furthermore, these factors are taken into account when establishing minimums, maximums, and zero points of the contingencies, since they are part of the historical data that are used.

An important example from the second group is the increase in malfunction rate due to poor quality of spare parts (including consumption materials such as toner and developer). This problem usually occurs on a specific range of copiers. Management explained that its hand were tied on the issue of second source materials, since the worldwide Gestetner organization determined which spare parts were to be used by its daughter organizations.

Excluding these uncontrollable factors would be difficult and might even have negative consequences. For example, excluding repeat calls caused by customer faults might cause a decrease of instruction of customers on the handling of photocopiers (which is something a technician is supposed to do if necessary). It would also not be feasible to adjust the contingencies each time a (temporary) problem occurs, even if the scope of the problem would be known. Moreover, immediate adjustment of the contingencies would undo the signaling function of the feedback. On the other hand, if changes in this area occur, they may result in adjustments to the contingencies, provided they are of a structural nature and their extent can be estimated.

3. Perceived overestimation of the 'cost'-product. A number of technicians argued that "a technician does a good job if his quality is high and, consequently, the customer is satisfied. The cost made are much less important. Therefore, the cost-product carries to much weight in the system; in this way, you risk getting dissatisfied customers."

In response to this and similar remarks–which perhaps reflect that most service technicians regard professional competency as attaining high quality rather than low costs–the facilitator again explained the procedure that was used in determining the relative importance of the indicators, in which both costs and customer satisfaction were used as criteria. Management explained that ProMES still reflected management policy in which cost control played an important part next to customer satisfaction. However, if in the future the focus were to shift toward customer satisfaction, the relative importance of the corresponding indicators could and would be easily changed accordingly[7].

4. Life cycle of photocopiers. The part of the life-span a certain type of copier is in determines to some extent the results that can be attained on the indicators. For example, new copiers require relatively little maintenance (including spare parts replacement) than older ones, which means that the cost of spare parts replacement per call increases during the lifetime of a copier. When this occurs, the contingencies should be adjusted accordingly.

[7] In the course of 1993, findings from a large scale survey concerning customer satisfaction caused extra activities aimed at increasing service-related customer satisfaction. This led to some changes in the ProMES system, such as a decreased weight of the cost-indicators and a broader definition of 'Labor time/call' which included contacting the key operator to discuss the repairs and maintenance carried out. Also, the development of a separate indicator for 'customer-satisfaction' was contemplated.

Facilitator and management agreed with the main points of these remarks, which again demonstrates that a regular check of the contingencies is in order. One should, however, not assume that the 'performance' of a population of a certain type of photocopiers deteriorates 'automatically', since machines that function poorly are usually the first ones to be discarded and replaced by a newer type. Therefore, the performance of the remaining copiers could very well remain up to the standard.

5. *Zero point.* Some technicians expressed their uneasiness about the use of negative effectiveness scores in the feedback report. They expressed their frustration with time and again receiving negative scores, which they felt were not an accurate reflection of their performance.

In response, the remarks made during the introductory meeting and the supervisor training program were repeated by the facilitator: it was stressed that the only meaning of the zero point was 'the average performance of all Nashuatec technicians in the Netherlands in some period in the past' and that its main function is providing the comparability of different indicators and types of copiers. This use of the zero point automatically results in a substantial part of the technicians receiving a negative overall effectiveness score. So, the zero point should not be considered a relevant point of reference for most technicians. For some technicians, an overall effectiveness of zero would constitute excellent performance, whereas for others it would be below the expected level. For example, technicians who recently started working for Nashuatec should not expect (or want to) achieve an overall effectiveness of zero; given sufficient ability, that score might be a goal for them after several months of experience at the job. Experienced technicians performing at a level far above zero should set their goals accordingly, since their level of ability and experience enables them to perform at a higher level of effectiveness. So, in general, taking past effectiveness as a starting point, technicians should think about ways of improving their performance–or in some instances: maintaining it. In this process, the zero point should be only of limited relevance.

In response, some technicians argued that the negative scores were largely the result of factors they could not influence. In some special cases this may have been true, for example if a technician was more or less based at a large company with a very old set of copiers. However, for the majority of technicians, no structural disadvantageous factors could be found. Taking into account that most comments were made by older, more experienced technicians, it seems mostly to be a matter of technicians having difficulty to accept that their level of performance is below 'average,. This is in accordance with the feedback literature which predicts lower acceptance of negative feedback due to lower perceived validity (Ilgen et al., 1979; see Section 2.4.2).

6. *Lack of information on causes and solutions.* A number of technicians mentioned that the link between their daily work behavior and the ProMES results was often unclear to them; ProMES feeds back the results of task behaviors, but neither shows the technicians the causes of these results nor what should be done to improve certain results. Given the issues pertaining to the zero point presented before, this caused some

frustration. In his response, the facilitator explained that ProMES is an output orientated system. ProMES typically feeds back the results of task behaviors and initiates discussion on causes and solutions. The contingencies and effectiveness scores obtained may direct attention to certain products and indicators and aid in setting priorities. However, the system does not prescribe certain task behaviors or strategies; these should result from a problem-solving discussion[8]. At Nashuatec this problem-solving discussion has been in the form of a periodical bilateral discussion between the supervisor and his individual technicians.

Part of the problem is caused by the service department policy which gives target time attainment (i.e. providing service within a certain time limit) a higher priority than sending the same technician to the same machine. Only in an estimated 50 percent of the calls does a technician return to a machine which he worked on the previous call. Only in those cases is he able to directly observe the consequences of his work on a tangible machine (e.g. what MCBC have I achieved, what was the cause of the malfunction, could I have prevented it, etc.). In case of a repeat call, target time attainment is especially critical, which means that technicians frequently do not have the opportunity to determine the cause of the repeat call and take corrective action themselves. Most regions recognized this problem and instituted a procedure for exchanging information on repeat calls between technicians.

Still lacking were information per type of copier, and information on malfunction causes, that could be useful input for technical meetings, conducted by the a senior-technician and focusing on a specific range of copiers. Management promised to look into possibilities for providing this information.

7. Long term effects/animosity between colleagues. In most regions, two risks associated with the ProMES program were mentioned. First, working with ProMES may cause a deterioration of long term quality because the technicians will be tempted to deviate from the maintenance procedure in order to save time and spare parts. This does not always immediately cause a decrease of short term quality as expressed by MCBC and percent repeat calls, and is therefore an advantageous short term strategy. However, the long term quality of the machine is compromised if these deviations become common practice. Secondly, the above strategy increases animosity between colleagues within a region, since parts of the maintenance not carried out cause a lot of extra work for the next technician on the machine (who in about 50 percent of the cases is someone else).

According to facilitator and management, the system includes 'group indicators' which should promote cooperation between technicians and secure the long term quality of the work done. The supervisor should perform his 'after visits' and 'visits

[8] In general, problem-solving is not limited to weekly or monthly feedback meetings. A lot of issues can be settled through on-the-job information-exchange and problem-solving. For example, a task group working with ProMES may discover that a strategy agreed upon during a feedback meeting is not effective, and decide to change it. When face-to-face interaction is not possible, such as in the Nashuatec situation, information-exchange and problem-solving is much more difficult.

during call' conscientiously and discourage technicians who break this agreement from maintaining this behavior. So, whether the risk mentioned becomes real depends to some extent on the supervisor. When asked whether they was willing and able to carry out their visits that way, the supervisors answered affirmatively. They, however, indicated that the checks performed only reveal serious deviations, not smaller ones. In addition to the supervisor, the group of technicians and senior-technicians also has a responsibility too insure that individual technicians remain accountable towards his colleagues in the region. To summarize, the question "should we 'polish or ProMES'?" was answered by "you should polish *and* ProMES". See Section 5.3.3 for the results on the group indicators. These results do not reflect wide-ranging neglect for long-term quality.

8. Use of ProMES in performance appraisal. By far the most controversial issue was the proposed use of ProMES as part of the performance appraisal instrument. In most regions, the majority of technicians were opposed to ProMES being used for performance appraisal purposes, although they did consider the system useful for regulating their own performance. When asked whether ProMES would provide better quantitative input about their 'technical' performance than sources used in the past, the majority responded affirmatively. Still, some technicians were outright opposed to using ProMES in the performance appraisal process, whereas others were opposed to the standardized procedure.

There seem to have been four causes for this predominantly negative attitude toward linking ProMES and the appraisal. First, some technicians did not understand the procedure that would be followed (although ample information on the subject had been provided at an earlier occasion). Erroneous assumptions by some technicians included: ProMES would make up 100 percent of the appraisal, using ProMES results lead to a lower mean bonus percentage than having the supervisor make up the appraisal, ProMES data of technicians who received almost no ProMES feedback would be used all the same, et cetera. Further explanation was necessary to clarify misunderstandings such as these.

A second likely cause is the increased uncertainty which the use of ProMES in the appraisal would entail. One aspect is the uncertainty about whether the supervisor is able and willing to take into account special circumstances (such as copy volume, low quality of some spare parts, problem-customers, etc.). Furthermore, due to the inability of the supervisors to 'justify' a relatively low bonus (except in very extreme cases) the variance in the bonus percentage paid in the past was very small since consequently no high bonuses could be paid due to a fixed total amount of bonus per region (in some regions, 90 percent of the technicians received exactly the same amount of bonus). This provides some certainty about which amount of money to expect, whereas the use of ProMES data would entail a larger dispersal of the bonus and, consequently, a less predictable situation. Still another aspect concerns a substantial category of technicians–especially the older, more experienced ones, some of whom had gotten used to a relatively high bonus percentage from the supervisor–almost by way of formality. If ProMES would be used, they would be confronted with a less positive result. Apart from this, the informal status of these technicians in the region might suffer somewhat. At an earlier occasion, this cause mentioned was by the supervisors, who observed that young, less experienced technicians appreciated the ProMES feedback very much, whereas a lot more experienced technicians rejected the feedback because it did not conform to their self-image.

The facilitator tried to take away the concern that the supervisor would not be able to take into account individual circumstances: the supervisor would still be determining 60 percent of the overall appraisal. A pilot appraisal in 'Enschede' and 'Utrecht' had demonstrated that the supervisor would have enough latitude if the situation called for it. Management added that it would expect the supervisor to take into account structural context differences, in addition to other responsibilities of the technicians, such as carrying out maintenance, filling in the history card, communication with clients and colleagues, et cetera. A standardized procedure was advocated because it would increase the fairness of the appraisal of regions, since supervisor would be obliged to look at the all important aspects of performance and apply the same standards and weights. In all regions, an example was given from the pilot appraisal, which involved a technician who had been receiving a relatively negative appraisal for a few years in a row, because of poor communication with clients. When ProMES was used in the appraisal it was found that this technician had been making an excellent 'technical' contribution (i.e. a very high overall effectiveness score in ProMES). Thus, in this instance, adding ProMES resulted in a more balanced appraisal by looking at both sides of the picture. Finally, the facilitator argued that giving all technicians the same bonus even though although clear performance differences were apparent (as was the case in the past) would be tantamount to saying that "it does not matter how you do your job", which would not be motivating technicians to do their work to the best of their abilities.

Thirdly, most senior technicians had a negative attitude towards the incorporation of ProMES into the appraisal. They indicated that they would prefer to be treated as a separate group in the performance appraisal, since their responsibilities differed significantly from those of the technicians. Although the current average overall effectiveness scores of technicians and seniors were almost identical (as shown by the facilitator), future changes in the seniors' job responsibilities could very well change this situation. After some further investigation into this issue, this request was granted by management.

Finally, both management and facilitator were reproached with having created the impression that each region would have complete freedom of choice regarding the use of ProMES as part of the performance appraisal. To some extent it was true that ambiguous signals had been given, although management and facilitator had added time and again that it would not be possible for the supervisor to disregard the information provided by ProMES, after a certain period of working intensively with the system. Although both acknowledged their lack of carefulness in this respect, this issue was not completely resolved. To some extent latter argument might have been used as a mock argument because this was the only argument that was used when most others were shown to be of limited validity.

In the next chapter (Section 6.5), the procedure for incorporating ProMES in the technicians' performance appraisal system will be discussed extensively, as will be some characteristics of performance management systems (such as ProMES) and performance appraisal systems that could explain the events that occurred in the Nashuatec project.

Discussion with the supervisors

After the evaluation meetings in the transportation regions, the supervisors from the participation and transportation regions discussed the implementation of ProMES with the facilitators and the field service managers. All in all, the supervisors considered ProMES a vast improvement over the information supplied in the past, and useful for getting a quick overview of the productivity of their regions and individual technicians and also as input in a performance appraisal. They, however, identified a handful of barriers to a successful use of ProMES.

First, the supervisors confirmed the technicians' feelings regarding the *zero point*. Even though the supervisors tried to draw away attention from the zero point and toward the technician's own effectiveness level, a lot of technicians who received predominantly negative feedback had trouble accepting this and expressed frustration at not being able to attain a zero level of effectiveness. In some cases, an explanation could be given for negative effectiveness scores (limited competency of the technician, insufficient effort, or an individual set of priorities that deviated from the set of priorities laid down in the ProMES system). However, in a number of cases, neither the technician nor the supervisor could find an explanation. Latter technicians became to view the feedback meetings as a negative experience.

Secondly, technicians and supervisors only had a limited insight into connections between the actions of a service technician and the results of these actions as depicted in ProMES. In other words: the *mechanisms* through which input is translated into output were *not clear* and, consequently, there were no clear strategies available for improving overall productivity. Moreover, the technicians expected the supervisors to provide them with quick-fix solutions, which in most cases was not possible. Also, the technicians were not accustomed to active participation in a problem-solving process. According to the supervisors, some additional information would be required to execute this problem-solving process effectively (e.g. ProMES information per type of copier, information on causes of repeat calls and MCBC-scores).

All supervisors agreed that the technicians considered ProMES to be a *performance appraisal system* as well as a performance management system. They also agreed that this posed a greater problem in the transportation condition than in the participation condition. In the participation condition, the technicians had been heavily involved in the design of the system, had participated in an effort to improve the accuracy and usefulness of the system, and had been involved in the process of devising a way of using ProMES in the performance appraisal that would be acceptable to them. According to the supervisors involved, this created a reasonably high level of confidence in the system in both its uses. In the transportation condition, on the other hand, problems such as an often limited understanding of the system, the limited controllability of some indicators, and the 'fakeability' of some elements of the system prevented full acceptance of ProMES as a performance management system (even though most technicians agreed that it was a more accurate measurement system than those used in the past). Use of ProMES for performance appraisal purposes would not be accepted, especially since the perceived opportunity to 'veto' a decision to use it for this purpose had been 'taken away'. The transportation supervisors feared that using ProMES in the performance appraisal would lead to competition between technicians (e.g. less attention to maintenance activities and cooperative behavior) and to (mostly undetectable) tampering with the system.

Finally, the bilateral feedback meetings both caused an *increase of the workload of the supervisors* and provided some *practical difficulties*. According to the supervisors, conscientious implementation of the feedback meetings would require some additional activities amounting to approximately 20 hours per month. In the past period this had not been achieved, due to the effort that had to be put into the attainment of the ISO certificate. The field service management representative argued that the bilateral feedback meetings merely represented a more structured way of carrying leadership activities that were an essential part of a supervisor's job. He did agree that management should attach priorities to the supervisors' activities. It was decided to start monitoring the allocation of time to the different activities that supervisors were required to carry out and to set priorities based on this information. Practical problems with the bilateral feedback meetings included finding time in the technician's work schedule for the feedback meeting and finding a suitable location for the meeting, Furthermore, some technicians lacked the discipline to have the necessary materials (e.g. their individual reports and graphs). The least time-consuming procedure appeared to be a pre-arranged meeting at the beginning of the day.

5.3. EVALUATION OF THE TRANSPORTATION PROCESS: EFFECTS ON PRODUCTIVITY

5.3.1. Research questions and method

Research questions

Three issues are central to the quantitative evaluation of the transportation phase of the project. First, the productivity effects of the transportation of ProMES on the productivity of the six regions involved are assessed. Secondly, possible additional effects on productivity due to bilateral feedback meetings in the participative condition are assessed. In Chapter 4, a substantial increase in productivity (18 points) was reported in the two regions that had taken part in the participative design of the system. However, the implementation of the system did not yet include the bilateral feedback meetings discussed in the preceding part of this chapter. In order to compare productivity changes in the participation and transportation conditions–which is the third central issue–the degree of additional productivity increase in the participation condition due to the bilateral feedback meetings should be determined. This results in the following questions:

1a) Is any productivity improvement obtained in the regions that did not participate in the design of ProMES (transportation condition)?

1b) Do the bilateral feedback meetings cause additional productivity improvements in the regions that participated in the design of the ProMES system (participative condition)?

1c) How does the productivity increase in the transportation condition compare to the overall increase in the participation condition?

ProMES productivity data

Productivity data, generated by the ProMES system, was used to assess the productivity change of individual technicians (their average productivity during a ten month period before and during a nine month period after the introduction of ProMES and the start of the bilateral feedback sessions). Three categories of technicians were excluded from the analysis. Because the ProMES measurement system covers only the approximately 30 copiers that are visited most frequently by the technicians, ProMES constitutes only partial measurement of the productivity of those technicians who mainly carry out maintenance on copiers not included in the measurement system, or other types of machines (such as fax-machines, laser printers, and copy printers)[9]. These technicians were excluded from the analyses[10]. A second category of technicians excluded from the analyses are technicians who entered the company's work-force

---

[9] During the feedback stage of the ProMES program, about 50 technicians were added to the field service department due to Nashua(tec)'s integration with another supplier of office machines. The work of these additional technicians differed from the work of the Nashuatec technicians in some respects. Two of these differences precluded meaningful productivity measurement and feedback for this group of technicians. Firstly, Nashuatec provides service on a relatively small number of different types of copiers that each generated a considerable workload. This enabled the development of a ProMES system which takes into account differences between types of copiers by using type-dependent indicators which have contingencies for each type of copier. In contrast, the additional technicians provide service on a wide variety of types of photocopiers, of which only a few types generate a workload comparable to the Nashuatec machines. This prevented the development of accurate contingencies. Secondly, whereas the Nashuatec technicians combine preventive maintenance with solving machine malfunctions ('opportunity service'), the additional technicians carried out preventive maintenance on separate occasions. Since the ProMES system was linked to the use of these opportunity service procedures, the ProMES system could not be applied to work procedures used by the new technicians.

[10] Included in the analysis were those technicians whose feedback reports contained at least 25 hours of labor time during at least 5 out of 10 months before the intervention and 5 out of 9 months after the intervention (both in the experimental and the control regions). This criterion was employed for three reasons: (1) the ProMES feedback is not relevant for technicians in the participation and transportation regions who receive feedback on a very small portion of their work, (2) random error due to a limited amount of data per month is reduced, and (3) consistency with the performance appraisal procedure (which requires 25 hours of labor time during at least 8 months per year for inclusion of ProMES data in the performance appraisal) is achieved.

shortly before or during the measurement period. These technicians go through a rather steep learning curve during their first few months 'in the field'. If included, this sub-category of technicians could pose a 'maturation threat' (e.g. Cook & Campbell, 1979) to the validity of the experimental approach taken here. Therefore, only technicians that entered the work-force at least three months before the start of pre-intervention measurement period were included in the analyses conducted. Thirdly, technicians who left the company during the measurement period were excluded from the analysis.

5.3.2. Individual indicators

Figure 5.2 depicts the ProMES effectiveness scores in the participation, transportation and control conditions, before and after the interventions (i.e. introduction of ProMES in the transportation condition and bilateral feedback meetings in both the participative and the transportation condition). Figure 5.3 gives a more simplified view of the before-and-after effectiveness scores.

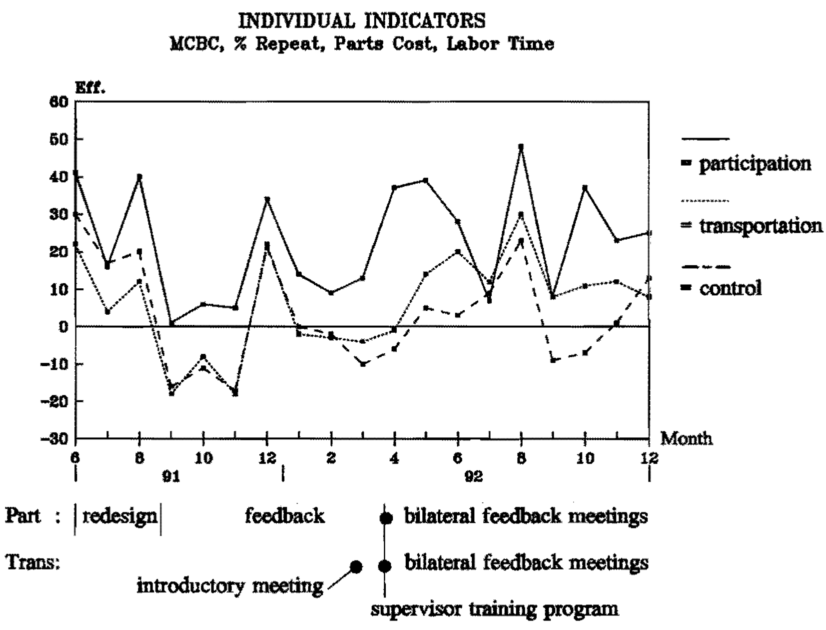


Figure 5.2. Individual indicators: effects.

Transportation condition

In six regions, the ProMES system was introduced by means of a standardized introductory meeting, which was followed by bilateral feedback meetings conducted by the region supervisors. These combined interventions resulted in a productivity increase of

11 ProMES points compared to the matched control regions. This difference was statistically significant (t_{150}=2.16, P=.017, one-tailed). A 2x2 MANOVA, in which the overall mean effectiveness from the periods before and after the introduction of ProMES were considered repeated measures, revealed a significant main time effect (before/after introduction of ProMES) and an interaction effect between condition (transportation versus control) and time (before/after introduction of ProMES). See Table 5.8 for analyses and results.

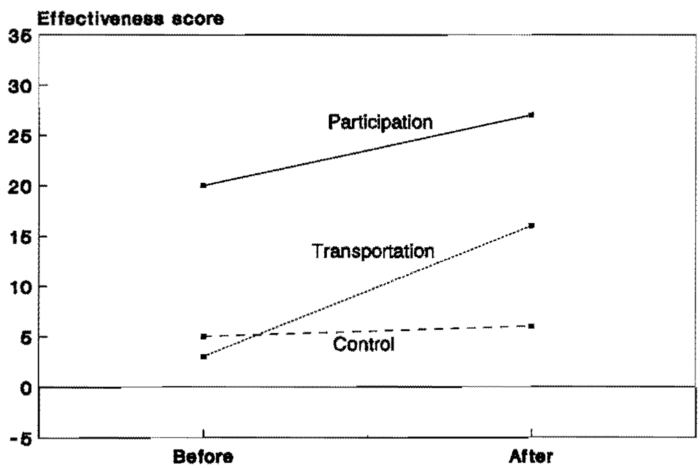


Figure 5.3. Individual indicators: summary of effects.

Table 5.8. Productivity changes: transportation versus control.

MANOVA, repeated measures

	SS	DF	MS	F	Sign.F
Between-Subjects Effects					
within cells	201350.05	150	1342.33		
condition (transport./control)	963.28	1	963.28	.72	.398
Within-subjects effects					
within cells	76764.94	150	511.77		
time (introd. + bil.feedback)	2947.82	1	2947.82	5.76	.018
condition by time	2381.90	1	2381.90	4.65	.033

Planned comparison of productivity change (t-test)

Condition	Number of Cases	Mean	Standard Deviation	Standard Error
Transportation	79	11.8	32.74	3.68
Control	73	0.6	31.16	3.65

t(150) = 2.16; P(one-tailed) = 0.017, significant

Participation condition

The regions that had participated in the design of the ProMES system achieved a significant productivity increase of 18 points after the system had been redesigned. The difference between the participation and the control condition increased to 25 points during the 9 month period in which bilateral feedback meetings took place. So, in this bilateral feedback-period, the *additional* average productivity increase in the participation condition amounted to 7 points compared to the control condition. A one-tailed t-test revealed that this additional increase was not statistically significant (t_{104}=1.16, P=.126). A 2x2 MANOVA showed a significant condition effect, but no significant time or interaction effect. More detailed information is presented in Table 5.9.

The fact that the additional productivity increase in the participation condition was not significant compared to the control condition should not lead to the conclusion that the bilateral feedback meetings were not successful for those who helped design the ProMES system. One might assume that it would be more difficult to keep improving an already high level of productivity (participation condition) than start improving a relatively low level of productivity (transportation condition).

Table 5.9. Productivity changes: participation versus control.

MANOVA, repeated measures

	SS	DF	MS	F	Sign.F
Between-Subjects Effects					
within cells	167007.49	104	1605.84		
condition (partic./control)	15497.64	1	15497.64	9.65	.002
Within-subjects effects					
within cells	48601.99	104	467.33		
time (bil. feedback meetings)	852.28	1	852.28	1.82	.180
condition by time	622.28	1	622.28	1.33	.251

Planned comparison of productivity change (t-test)

Condition	Number of Cases	Mean	Standard Deviation	Standard Error
Participation	33	8.0	29.21	5.08
Control	73	0.6	31.16	3.65

t(104) = 1.15; P(one-tailed) = 0.126, not significant

Effects per indicator

Table 5.10 presents the mean absolute effectiveness levels of the three conditions in the bilateral feedback period. In the previous chapter, it was shown that the 18 point increase in overall effectiveness that resulted from implementing the re-designed ProMES system in the regions of 'Enschede' and 'Utrecht' (the participative condition) was mainly due to improvements on the cost-indicators ('Labor time/call': +9, 'Parts cost/call': +5). It can be gathered from Table 5.11 that the additional increase in the participation condition was entirely brought about by increased effectiveness on the *cost* indicators–especially 'Labor time per call'. In contrast, the increase in the transportation condition was almost entirely the result of increased effectiveness on the *quality* indicators. Thus, on average, the two conditions receiving ProMES feedback seem to have used a different (implicit) strategy.

Table 5.10. Effectiveness scores per indicator (bilateral feedback period).

Indicator	part	trans	contr
Mean Copies Betw. Calls	–5	–7	–12
Percent repeat calls	11	16	13
Parts cost/call	2	1	1
Labor time/call	19	5	4
Overall effectiveness	27	16	6

Table 5.11. Effectiveness change per indicator (before/after introduction and start bilateral feedback).

Indicator	part	trans	contr	part vs contr	trans vs contr	combined change participation
Mean Copies B. Calls	-7	-1	-7	0	6 *	3 + 0 = 3
Percent Repeat calls	5	11	7	-2	4 *	1 –2 = –1
Parts cost/call	0	0	-2	2	2	5 + 2 = 7
Labor time/call	10	2	3	7 *	-1	9 + 7 = 16
Overall effect. change	8	12	1	7	11 *	18 + 7 = 25

* = P<.05 (paired t-test, df=8)

Effects per region

As is revealed in Table 5.12, the two regions in the participation condition attained a comparable effectiveness increase (7 and 8 points respectively). However, there are large differences between the transportation regions, ranging from a 26-point increase in region Haarlem to a 10-point decrease in region Eindhoven. One possible reason for these differences–the 'quality' of the bilateral feedback sessions–will be examined in the next chapter. The effectiveness changes in the control conditions range from a 11-point increase to an 16-point decrease.

Table 5.12. Productivity effects per region

Condition/regions	before	after	change
PARTICIPATION (mean)	20	28	+8
2: Enschede	23	30	+7
3: Utrecht	18	26	+8
TRANSPORTATION (mean)	3	15	+12
4: Arnhem	25	43	+18
6: Eindhoven	19	9	−10
7: Maastricht	−9	4	+13
8: Haarlem	−9	17	+26
11: Den Haag	−1	6	+7
12: "Hoevelaken"	−1	16	+17
CONTROL (mean)	5	6	+ 1

Effects of 'gaming the system'

According to Pritchard (1990, pp. 154-155), any feedback system can be 'gamed'. That is, the unit personnel can find a way to distort the information in order to make themselves look good. ProMES tries to minimize this tendency by having unit personnel participate in the development of the system. If unit personnel perceive the system as 'their system' and as a valid one, they will be less inclined to distort it. It may, however, still be desirable to carry out some outside checks in order to detect distortions. This holds especially true if the unit itself is collecting and reporting the data that will be the basis of their own feedback reports.

For the ProMES system at Nashuatec two issues are relevant when considering 'gaming of the system'. The first issue pertains to the lack of participation in the transportation condition, the negative attitude of some technicians toward the validity of the system, and, in particular, toward the likelihood of future use of the ProMES system as part of the performance appraisal system. This complex of factors might have prompted technicians to distort some of the data to 'look good in the figures'. In the second place, data collection on the individual indicators is done by the technicians themselves. The company's information system features some built-in checks to detect unlikely data entrances (e.g., a reported copy counter which is lower than the one reported the previous call). However, technicians do have the opportunity to make their results look slightly better than they really are.

A distortion which is difficult to detect at the level of individual technicians but which can be detected at the level of the three conditions defined in this thesis concerns the indicator 'Labor time per call'. The technicians supply data on different categories of time-use on the job (e.g. labor time, travel time, deliberation with product support, contact with the customer's key operator). To ascertain whether the decreased amount of labor time used is not in fact the result of an administrative shift of labor time to travel time or to the other categories of time used–both of which are not included in ProMES–two checks were performed.

The first check involved a comparison of changes in the travel time per call for the ProMES and control conditions. During the post-redesign feedback period (discussed in the previous chapter) no shift from travel time to labor time occurred in the participation condition. In the next phase (discussed in this chapter), there seems to have been a small shift from labor to travel time in the transportation condition, which amounts to approximately 1½ minute per call. In the participation condition, this shift amounted to 1 minute per call.

The second check involves a comparison of changes in the use of special categories of time-use in the ProMES and control conditions. Again, no changes occurred in the post-revision period in the participation condition. In the next phase, both the participation and transportation increased the use of 'special time' by approximately 1 minute compared to the control condition.

Using the contingencies, the above units of time can be converted into ProMES effectiveness: one minute equals 0.8 effectiveness points. The effectiveness score of labor time per call would decrease 2 points in the transportation condition and 1½ points in the participation condition[11].

In summary, these differences are very small and may not even be the result of consciously 'fraudulent' use of the ProMES system by some technicians; it is even possible that these differences result from a more *accurate* use of the time-codes available by the technicians working with ProMES.

5.3.3. Group indicators

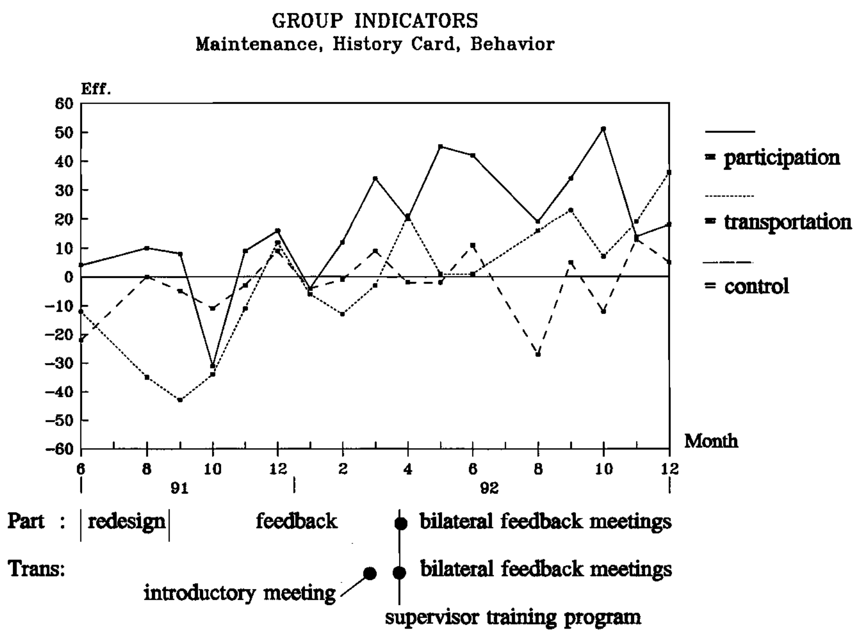
The main focus of the ProMES system is on individual productivity, expressed by the four individual indicators. The group indicators, measured by the region supervisor, primarily serve as a safeguard against negative long-term effects of maximizing short-term effectiveness. For example, some technicians expressed the fear that technicians who were trying to improve their productivity by reducing costs (labor time per call and/or parts cost per call), would do so by sloppily carrying out the maintenance procedure, not adhering to prescriptions for preventive replacement of spare parts, and not filling in the History Card, among others. If these occurrences would have been frequent, this would be reflected in a lower group level effectiveness score (particularly for the preventive-maintenance indicator). Supervisors confirm that a slight deviation from the procedure would probably be difficult to observe. However, technicians who

[11] If these two points are deducted from the overall effectiveness change in the transportation condition, the difference with the control condition is reduced to 9 points. This difference remains significant (t[152, one-tailed]= .030).

deviate frequently and extensively from the prescriptions are almost certainly noticed within a few months.

Figure 5.4 shows the monthly group effectiveness values for the three conditions. The group effectiveness score in the control condition has been relatively stable throughout the pre- and post-introduction period. Although direct comparison of measurements done by different supervisors should be dealt with cautiously due to possible small differences and changes over time in standards used, some inferences can be made on the basis of overall results per condition.

Figure 5.5 shows a marked increase in the transportation condition's effectiveness from approximately 15 points below the control condition in the pre-introduction period to approximately 15 points above the control condition in the post-introduction period. There has also been an additional increase in the participation condition of about 25 points.



Note: because of the limited amount of data collected in the holiday period (July and August), results of these months are combined and shown under August.

Figure 5.4. Group indicators: effects.

The overall results for the post-introduction period are presented in Table 5.13. The overall effectiveness of the participation and control condition is significantly higher than that of the control condition. The difference between the participation and transportation condition is not significant.

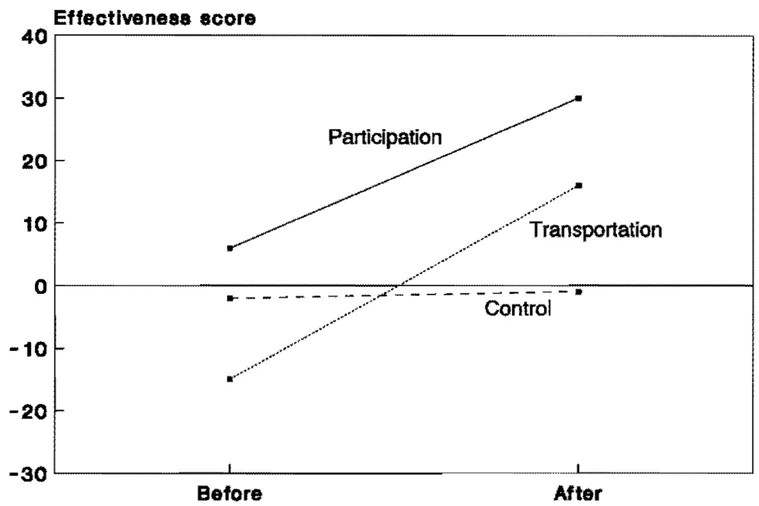


Figure 5.5. Group indicators: summary of effects.

When calculating the effectiveness change per indicator (Table 5.14), a large increase in the transportation condition on the preventive maintenance indicator becomes evident. This is consistent with the 'quality-oriented'-strategy which seems to have been dominantly followed in these regions. The participation condition also shows a small increase on this indicator, compared to the control condition. This is not in accordance with the contention of some technicians that a 'cost-oriented strategy' causes neglect of long-term quality, and therefore offers some evidence that this is not a widespread practice in the regions concerned.

Table 5.13. Group indicators: mean effectiveness per indicator (bilateral feedback period).

	part	trans	contr	part vs contr	trans vs contr	part vs trans
Preventive maintenance	2	-5	-10	12 *	5	7
History Card	2	1	–3	5	4	1
Behavior	26	19	12	14 *	7 *	7 *
Overall group effectiveness	30	16	–1	31 *	16 *	15

* = P<.05 (one-tailed if involving comparison with control condition, others two-tailed)

Table 5.14. Effectiveness change per indicator (before versus after introduction/start bilateral feedback).

	part	trans	contr
Preventive maintenance	+ 8	+ 19	+ 3
History Card	+ 6	+ 5	+ 1
Behavior	+ 10	+ 7	- 3
Overall group effectiveness	+ 24	+ 31	+ 1

5.4. FINANCIAL EFFECTS OF IMPLEMENTING PROMES

5.4.1. Introduction

In order determine the financial effects of the implementation of ProMES so far, the results of the Activity Based Costing analysis–used to help establish the relative importance of the indicators (Section 4.6.6)–were applied to the productivity changes reported in this chapter and the previous one.

Taking into account the fact that ProMES covers approximately 75 percent of the workload of the field service department, the potential monetary value of one ProMES effectiveness-unit can be calculated. Per indicator, a nation-wide effectiveness increase of one point per month corresponds with the following (potential) cost-reduction:

Mean Copies Between Calls	Dfl 4.125	(Dfl 1.00 = approx. $ 0.55)
Repeat Calls	Dfl 1.725	
Parts Cost per call	Dfl 7.800	
Labor Time per call	Dfl 4.425	

The above amounts solely refer to the potential cost reduction and not to increased customer satisfaction which may result from increased effectiveness on the indicators. An estimate of latter non-quantifyable gain was included in the procedure for determining the relative importance of the indicators. Therefore, the gain per ProMES point is not equal for all four indicators, which it should have been if cost reduction had been the only result of improved effectiveness.

5.4.2. Establishing the potential cost reduction

The potential cost reduction is established by the effectiveness change in the participation and transportation groups to that of the control groups. The overall effect is divided into three sub-effects:

1. Post-redesign feedback period in the participative condition (June 1991–March 1992).
2a. Bilateral feedback period in the participative condition (April–December 1992).
2b. Bilateral feedback period in the transportation condition (April–December 1992).

Results are shown in Table 5.15.

Table 5.15. Potential cost reduction per condition per period

1. Post-redesign feedback period in the participative condition (June 1991–March 1992), 18 percent of the technicians; 10 months.			
indicator	eff. change	potential cost reduction	
Mean Copies Betw. Calls	3	3 * 4125 * 10 months * 18 %	22,280
Repeat Calls	1	1 * 1725 * 10 months * 18 %	3,105
Parts cost per call	5	5 * 7800 * 10 months * 18 %	70,200
Labor time per call	9	9 * 4425 * 10 months * 18 %	71,685
Overall	18		Dfl 167,270

2a. Bilateral feedback period in the participative condition (April 1992–December 1992), 17 percent of the technicians; 9 months.			
indicator	eff. change	potential cost reduction	
Mean Copies Betw. Calls	3	3 * 4125 * 9 months * 17 %	18,934
Repeat Calls	−1	−1 * 1725 * 9 months * 17 %	−1,725
Parts cost per call	7	7 * 7800 * 9 months * 17 %	83,538
Labor time per call	16	16 * 4425 * 9 months * 17 %	108,324
Overall	25		Dfl 209,071

2b. Bilateral feedback period in the transportation condition (April 1992–December 1992), 39 percent of the technicians; 9 months.			
indicator	eff. change	potential cost reduction	
Mean Copies Betw. Calls	6	6 * 4125 * 9 months * 39 %	98,010
Repeat Calls	4	4 * 1725 * 9 months * 39 %	27,324
Parts cost per call	2	2 * 7800 * 9 months * 39 %	61,776
Labor time per call	−1	−1 * 4425 * 9 months * 39 %	−17,523
Overall	11		Dfl 169,597

Total (potential) cost reduction per indicator	
indicator	cost reduction
Mean Copies Between Calls	139,224
Repeat Calls	28,704
Parts cost per call	215,514
Labor time per call	162,486
Overall	Dfl 545,928

A comparison of the overall (potential) cost reduction and an estimate of the expenses involved in the designing, implementing, and transporting the system (Table 5.16) suggests that the 'break-even point' would be reached if 54 percent of the potential cost-reduction were indeed realized.

Table 5.16. Estimated expenses (meetings, facilitators, training program)

Condition/phase		Expenses (estimates)
Participation condition:	design	Dfl 96,000
	implementation	Dfl 96,000
Transportation condition		Dfl 100,000
Overall expenses		Dfl 292,000

The following example illustrates the potential cost reduction of a full-scale implementation of ProMES. If the remaining six regions were to attain the same effectiveness increase as the transportation condition (11 points), the implementation of ProMES would result in an average nation-wide effectiveness increase of 14 points (MCBC: +5, percent repeat: +3, parts cost: +3, labor time: +3). This corresponds with a potential cost reduction of approximately Dfl 750,000 in one year.

5.4.3. Potential versus actual cost reduction

Some preconditions should be fulfilled for the potential cost reductions to be converted into actual cost reductions. All indicators, with the exception of 'Parts cost per call', cause a reduction in cost primarily through reduced use of labor (if technicians succeed in realizing a shorter labor time per call, the total amount of labor time needed decreases; if the number of repeat calls diminishes or the number of copies between successive calls increases, the total number of calls per period decreases). These potential gains will equal a zero actual gain if the time won is idle time (e.g. if technicians are sent home half an hour early each day in stead of carrying out an extra call). If the organization decides to do the work with a smaller work-force (for example by natural turnover), an actual cost reduction is achieved. An alternative toward transforming potential gains into long-term gains is using the extra time for carrying out activities aimed at increasing customer satisfaction. In deciding which alternative should be pursued, the estimated gains from increased customer satisfaction should be weighed against the actual cost reductions resulting from a decreased work-force. Parts cost per call is probably the only indicator which requires no conscious management decision for potential cost reductions to materialize.

5.5. COMPARISON OF THE PARTICIPATION AND TRANSPORTATION EFFECTS

Both in the participation and transportation condition, the implementation of ProMES caused a significant productivity increase on the individual indicators compared to the control condition. However, a direct comparison of the effects on productivity of implementing ProMES in the participation and control condition is hindered by differences in the length of the intervention periods. The participation regions achieved a mean increase of 21 points in a 19 month period[12] (excluding the 7 month pilot feedback period in which the system had to be extensively redesigned). The mean productivity increase in the transportation regions was 11 points during a 10 month period. So, it could be argued that the increase per month is about equal for both conditions (about two points). Nevertheless, there are some indications that this line of reasoning–which presupposes a gradual continuous increase in productivity–does not hold true. Judging from Figures 4.8 and 5.2, the interventions did not result in a gradual productivity increase. Rather, these figures typically show a sudden increase in the first few months of the intervention period; this increased level is maintained but not much improved in the later months of the intervention period. Judging from Figure 5.2, the transportation regions may not have been able to maintain their productivity increase: in the final month of the intervention periods, their productivity equals the productivity of the control group. Finally, formalizing the feedback by instituting bilateral feedback meetings caused an additional seven-point increase in the participation condition, which is rather high, taking into account the already relatively high level of productivity in the participation condition and the 11-point increase in the transportation condition during the same period. In all, the data support the following conclusion: *the productivity increase in the transportation regions is approximately half that of the participation regions.*

This conclusion is confirmed by the results of an effect-size calculation (*d* statistic), which was also used by Pritchard et al. (1988, 1989) (see Section 1.4.2). Pritchard et al. (1988, 1989) calculated the *d*-statistic by dividing the mean difference between baseline and treatment periods by the pooled within-groups standard deviation in those periods. In this study, the pilot feedback period is regarded as the baseline period. Also, control group effectiveness is taken as a reference point for the effects in the experimental conditions [Eff'(exp)= Eff(exp)–Eff.(control)]. As in Pritchard et al. (1988, p. 349), a conservative value for the combined effects over units is used (i.e. first calculating the effect size per unit, than taking the mean across units). The results, shown in Table 5.17, indicate that the participation effect is approximately 2.5 times as large as the transportation effect.

---

[12] That is an 18-point difference during the ten-month post-redesign period and an 25-point difference during the nine-month bilateral feedback period: [(+18) * 10 + (+25) * 9]/19 = +21.

Table 5.17. Comparison of effects in the participation and transportation condition (d statistic)

Condition	Effect	Feedback
Participation (2 regions)	Mean	2.08
	Range	1.62 (Utrecht) - 2.53 (Enschede)
Transportation (6 regions)	Mean	0.82
	Range	−0.80 (Eindhoven) - 1.74 (Haarlem)

Both in the participation and the transportation condition, group indicator effectiveness increased as a result of the implementation of ProMES. This seems to indicate that the increase of short-term effectiveness as measured by the individual indicators has not harmed long-term quality or cooperation between technicians.

Attitudes of participation and transportation technicians toward the system will be discussed in the next chapter.

5.6. CONCLUDING REMARKS

In the previous chapter, it was demonstrated that it was possible to design a ProMES control loop in a setting which possessed five main complicating characteristics (interdependence between individual workers, a complex work flow, a lack of communication, a top-down culture, and concern for performance appraisal and reward issues). After redesign of the control loop, in order to conform more closely to the design criteria, positive effects on productivity were obtained. In this chapter, it was demonstrated that it was possible to transport the system to other groups, without going through the design process. However, the mean productivity improvement obtained in this non-participative condition was less than half of the improvement obtained in the regions that participated in the design of the system.

In the next chapter, we will examine to what extent the criteria for an effective control loop–set out in Chapter 2–were met: Do the feedback and goal mechanisms operate optimally? To what extent did the approach taken to the design process and the implementation process confirm to the requirements? Does the design itself meet the requirements? Were the situational constraints posed by the organizational context handled adequately? What could have been done (or: could be done) to improve the effectiveness of the control loop? These questions are the basis for a qualitative assessment of the effectiveness of the ProMES control loop at Nashuatec.

Chapter 6
Evaluation

6.1. INTRODUCTION

This chapter is subdivided into two parts. In the first part of the chapter, an evaluation of the effectiveness of the control-loop design will be given, based on the design criteria and relevant situational constraints described in Section 2.4 and summarized in Figure 6.1. We will examine to what extent the general design criteria for feedback and goals have been met, to what extent requirements from the ProMES approach have been fulfilled, and how effectively situational constraints have been taken into account. This evaluation serves two purposes. In the first place, it can give an indication as to which aspects of the approach taken have been successful and which have not produced the desired effects and, consequently, what should be done to further enhance the effectiveness of the control loop. Secondly, it serves to expand the 'pragmatical design-knowledge base' on factors that should be taken into account when designing and implementing (ProMES-) performance management systems. This can aid in producing more carefully balanced and, thus, (even) more successful projects.

The second part of this chapter focuses on some interesting by-products of the ProMES program, such as the connection that was made between ProMES and the performance appraisal of the technicians and the 'management by results'-system of the supervisors in the field service department.

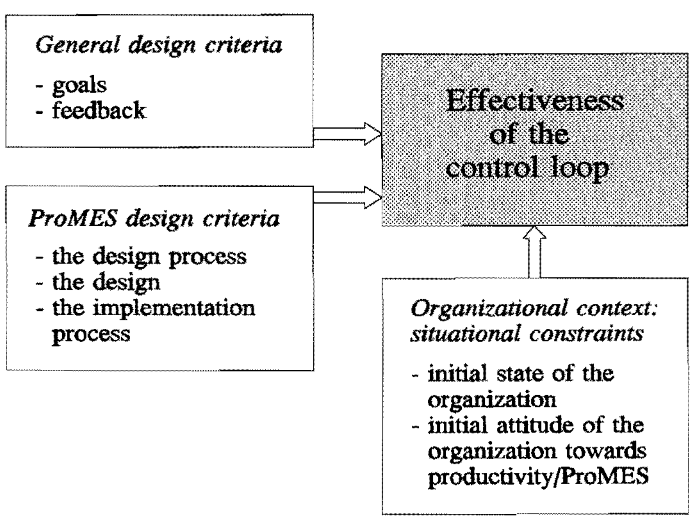


Figure 6.1. Design criteria and situational constraints.

6.2. ATTAINMENT OF GENERAL DESIGN CRITERIA

6.2.1. Introduction

One would hypothesize that the implementation of ProMES in a group would provide a control loop which meets the general design criteria for effective goal setting and feedback (discussed in Section 2.4.2) to a larger extent than control loops that have existed in that group before the implementation of ProMES (or: that exist in comparable groups not using ProMES at the same time). This results in the following research question:

"Does the implementation of ProMES improve the extent to which the general design criteria for effective feedback and goal setting are met?"

This question involves the comparison of the participation, transportation, and control conditions, using a questionnaire which measures to what extent the criteria for effective feedback and goal setting are fulfilled. Successively, we will discuss the design used to answer this question, the specific questionnaire used, and the main findings.

6.2.2. Design

In all regions, a questionnaire was filled in approximately eight months after the introductory meeting and the start of the bilateral feedback meetings. This yielded posttest data on design criteria fulfillment which are the basis for the main results presented in this section.

In case only posttest data are available–which represents a 'posttest-only design with non-equivalent groups' (Cook & Campbell, 1979)–selection differences between the different groups are a possible threat to accurate assessment of treatment effects. To assess whether it would be safe to draw conclusions about the effects of the ProMES interventions based on post-intervention observations only, a limited pretest was done: the questionnaire was filled out in six of the twelve control regions before the transportation of ProMES started[1]. Two comparisons were made based on these data. First, for technicians in the control condition of whom both pretest and posttest were available, the pretest was compared with the posttest. Secondly, the pretests for the future control and participation conditions were compared. In both instances, only small and non-significant differences were found. This means that any posttest differences to be found between the three conditions may be attributed to interventions connected with ProMES.

[1] Originally, it was intended to use a Solomon Four Group design. However, time constraints and difficulties with matching the four groups on relevant variables prevented effective use of this design.

6.2.3. The "Goal setting and Feedback Questionnaire" (GFQ)

A goal setting and feedback questionnaire was used to measure, among others, the extent to which criteria for effective feedback and goal setting were fulfilled in the field service department. The fulfillment of these criteria is considered an essential precondition for substantial productivity improvement. The GFQ, adapted from a goal setting and feedback questionnaire developed and tested by Van Tuijl and his colleagues (see Algera & Van Tuijl, 1990), contains 77 items in 12 sections. The sections (number of items) are:

1. setting goals (8),
2. goal difficulty and attainability (8),
3. goal attainment strategies (4),
4. knowledge of priorities (4),
5. controllability of results (10),
6. dependence (2),
7. influence (2),
8. measurement of results (6),
9. feedback - amount and sources (10),
10. feedback - usefulness (10),
11. reinforcement (6)
12. relationship with supervisor (7)[2]

Each question in the first 11 sections was formulated as a statement. Subjects were asked to respond to these statements by giving a rating on a five-point 'Likert' scale added to each statement. The last section contained seven five-point semantic differential items. A few typical examples of items used are (translated from Dutch):

6.1	My performance depends on how my colleagues in the region do their job.	strongly disagree	1	2	3	4	5	strongly agree
11.1	If my performance meets the requirements, my supervisor lets me know that this is appreciated.	almost never	1	2	3	4	5	almost always
12.6	My relationship with my supervisor is...	tense	1	2	3	4	5	relaxed

Eight months after the start of the bilateral feedback sessions, the GFQ was distributed among all 283 service technicians then working for Nashuatec; the technicians individually filled out the questionnaire during a region meeting. The response was 89 percent (253 technicians). Data on the reliability of the scales used is included in Appendix D.

[2] Adapted from Nathan et al. (1991), who found that the ongoing interpersonal relationship between supervisor and subordinate influences the effectiveness of performance appraisal interviews. Since the bilateral ProMES meetings bear some resemblance to performance appraisal sessions, the relationship between supervisor and technician is included as a context variable.

6.2.4. Main results

The main results from the GFQ are divided into three categories: feedback criteria, goal criteria, and context factors (see Section 2.4 for an illustration of these categories). The basis for discussion of each category is a table which contains the most important scales (or single items) in the category, the mean scores from the three conditions, and an indication of any significant differences between the conditions[3].

Feedback criteria

In Table 6.1, information is given about the amount, utility, sign and source of feedback.

Amount of feedback. As could be expected, the implementation of ProMES caused a large increase in the perceived amount of feedback on results provided to technicians. In addition, also an increase–although smaller–occurred regarding feedback about work strategies. These findings may reflect the difference in attention for individual and group indicators in the ProMES system.

Feedback utility. The score on feedback utility encompasses five items: timeliness, completeness, availability, understandability, and specificity (related to goal attainment and to corrective strategies). These items reflect the feedback criteria listed in Section 2.4.2 and may be considered indicative of the extent to which feedback is accepted by persons who are to use it for regulating their performance (together with the sign of feedback, which is discussed next).

The three conditions differ significantly with regard to perceived feedback utility. Both in the participation and in the transportation condition, perceived feedback utility is higher than in the control condition. Interestingly, the participation condition considers the feedback significantly more useful than the transportation condition. This difference in utility is reflected by higher scores on perceived completeness, understandability and specificity in the participation group, whereas the scores on timeliness and availability are comparable for both conditions (see Table 6.2). The former factors all concern the contents of the feedback. These results are accordance with comments made in the transportation regions during the evaluation meetings, the majority of which reflect technicians' reservations about the contents of the feedback report.

Sign of feedback. The introduction of ProMES resulted in a small increase of positive feedback compared to the control condition. However, the amount of negative feedback

[3] Unless noted otherwise, t-tests involving a comparison between an experimental condition (participation or transportation) and the control condition are one-tailed (the hypothesis being that implementing ProMES leads to better fulfillment of the design criteria). T-tests involving a comparison between the experimental conditions are two-tailed.

did not change. A comparison of the amount of positive and negative feedback received revealed that significantly more negative than positive feedback was received in the transportation and control conditions (P=.001 and .044 respectively, using a paired t-test), whereas there was no difference in the participation condition. Keeping in mind the problems associated with the 'zero point' which were brought forward during the evaluation meeting, the result from the transportation condition was not surprising, although it also means that the supervisors might not have "stressed the positive" as much as would have been desirable.

Table 6.1. GFQ: feedback criteria.

Scale/item	(1) part	(2) trans	(3) cont	(4) one-way	(5) part vs trans	(6) part vs cont	(7) trans vs cont	(8) exp vs cont	direction of effect
Amount of feedback about results	4.23	4.01	2.41	.000	ns	.000	.000	.000	(P=T)>C
Amount of feedback about work strategies	3.28	3.27	2.67	.002	ns	.005	.001	.000	(P=T)>C
Feedback utility (see Table 6.2)	3.59	3.19	2.48	.000	.012	.000	.000		P>T>C
Negative feedback	2.93	2.85	2.97	ns				ns	
Positive feedback	2.76	2.49	2.31	ns				.048	PT>C
Feedback from regular reports	3.07	2.96	1.91	.000	ns	.000	.000	.000	(P=T)>C
Written feedback	4.00	3.82	2.31	.000	ns	.000	.000	.000	(P=T)>C
Feedback from supervisor	3.28	3.78	3.44	.041	.024	ns	.028		T>(C=P)
Verbal feedback	2.14	2.53	3.07	.000	ns	.000	.005	.000	(P=T)<C

ns = not significant (P>.05)
Explanation of columns: (1) average score participation condition (n=29); (2) average score transportation condition (n=79); (3) average score control condition (n=67); (4) multiple comparison test (oneway analysis of variance); (5)(6)(7) t-test for differences between conditions, provided the multiple comparison test is significant; (8) t-test for difference between control condition and combined participation and transportation condition (n=108), provided participation and transportation do not differ significantly; (9) formula, expressing the direction of the effect.

Source of feedback. As could be expected, the amount of written feedback received is significantly higher in the conditions working with ProMES. Specifically, the amount of feedback from regular reports is rated much higher in the experimental conditions than in the control condition. The supervisor as source of feedback has increased somewhat in the transportation condition, probably due to the start of the bilateral feedback meetings.

Table 6.2. GFQ: feedback utility, comparing participation and transportation.

Feedback utility	part	trans	difference	(sign. level)
specificity (goal attainment)	3.83	3.24	+ 0.59	(.001)
specificity (corrective strategies)	3.72	3.22	+ 0.50	(.024)
timeliness	3.00	3.05	- 0.05	(n.s.)
availability	3.21	3.08	+ 0.13	(n.s.)
completeness	3.79	3.31	+ 0.48	(.020)
understandability	4.00	3.27	+ 0.73	(.002)
Overall	3.59	3.19	+ 0.40	(.012)

n.s. = not significant (P>.05)

Goal criteria

Design criteria for goals include goal specificity, and goal difficulty and attainability. These are included in Table 6.3, along with results on participation in goal setting, controllability of results, and knowledge of strategies and priorities.

Goal specificity. Contrary to expectations, perceived goal specificity in the experimental condition does not differ from the control condition, both for desired results and work strategies.

The lack of difference may partly be explained by a difference of interpretation: technicians in the control condition may have interpreted goal specificity as guidelines from the maintenance procedure which they have to adhere to (e.g. replace this spare part after 50,000 copies, first clean this unit, than that one, report to the key operator when finished, etc.). This seems to be confirmed by the answers given regarding the technicians' knowledge of strategies toward goal attainment. In the control condition, the perceived knowledge of these strategies is higher than in the experimental conditions. Assuming that the maintenance procedure is equally well known in all conditions, this suggest a difference in interpretation of the concept of 'goals'. However, results from another questionnaire aimed at the bilateral feedback meetings (which will be discussed later on in this chapter) indicate that very few of these meeting resulted in the setting of specific quantitative goals, although frequently 'behavioral' agreements were made. This is not surprising, since the setting of specific goals constitutes the last phase of the feedback meetings and requires that the previous phases (agreement on the procedure, understanding of the feedback report, agreement on actions that lead to productivity-improvement) be concluded.

In all, it seems that one of the objectives of the feedback training–using ProMES to set specific productivity goals–was not attained.

Goal difficulty and attainability. In the control condition, goals are considered rather easily attainable, judging from a rather low mean score of 2.06 on the goal difficulty and a rather high score of 3.85 on the goal attainability scale. Technicians in the experimental conditions consider the goals somewhat more difficult and less easily attainable, although the difference is not very large. So, goals are still not considered very challenging, which diminishes the effectiveness of goals set.

Participation in goal setting. Although the feedback training emphasized a participative approach to feedback meetings, no significant differences were found in this respect.

Table 6.3. GFQ: goal criteria.

Scale/item	part	trans	cont	one-way	part vs trans	part vs contr	trans vs contr	exp vs contr	direction of effect
Goal specificity (results)	3.06	3.10	3.12	ns				ns	
Goal specificity (strategies)	3.97	3.99	3.84	ns				ns	
Goal difficulty	2.31	2.57	2.06	.002	ns	ns	.000	.001	T>C
Goal attainability	3.44	3.40	3.85	.001	ns	.006	.001	.000	(P=T)<C
Participation in goal setting	3.24	3.39	3.45	ns				ns	
Controllability of results	2.67	2.56	2.90	.008	ns	ns	.003	.005	T<C
Knowledge of strategies	3.29	3.26	3.73	.001	ns	.002	.001	.000	(P=T)<C
Knowledge of priorities	3.93	4.07	4.01	ns				ns	

ns = not significant at .05 level

Controllability of results. Since very few quantitative goals have been set, the results regarding this criterion are probably more an expression of the general feeling of the technicians that a lot of factors influence their results (which is confirmed by a mean score of 2.72). For the experimental groups the results may reflect an assessment of the controllability of the indicators. The slightly lower perceived controllability of results in the transportation condition compared to the control condition may be caused by interpretation differences.

Across all conditions, the results reflect the tendency for persons to attribute positive outcomes primarily to their own knowledge, skills and effort (mean score: 3.54) and to attribute negative outcomes primarily to external factors and not to a lack of knowledge, skills and effort on their part (mean score: 1.90).

Context factors

Table 6.4 presents results pertaining to some context factors that may influence the effectiveness of the feedback and goal setting program.

Possession of knowledge and skills. Both in the experimental and control conditions, the technicians report that they possess sufficient knowledge and skills to meet the performance requirements. Therefore, in general, goal acceptance is not compromised by a perceived lack of knowledge or skills.

Availability of means needed. Technicians in all conditions report that means needed to meet performance requirements are available (scores around 4.0). Technicians in the experimental conditions are a little less positive than their colleagues in the control condition. The reported lack of information on causes and solutions (one of the topics during the evaluation meetings) may explain this small difference.

Table 6.4. GFQ: context factors.

Scale/item	part	trans	contr	one-way	part vs trans	part vs cont	trans vs cont	exp vs contr	direction of effect
Possession of knowledge/ skills	4.03	4.19	4.33	ns				ns	
Availability of means needed	3.83	3.91	4.16	ns				.021	PT<C
Dependence on colleagues	3.24	3.32	2.95	ns				ns	
Dependence on supervisor	2.24	2.30	2.34	ns				ns	
Influence on colleagues	3.28	3.32	3.37	ns				ns	
Influence on supervisor	2.96	3.18	3.21	ns				ns	
Relationship with supervisor	4.43	4.24	4.22	ns				ns	

ns = not significant at .05 level (two-tailed)

Dependence and influence. One of the context factors which had been considered in designing the ProMES system is the interdependence between technicians. The dependence as perceived by the technicians was moderate (scores around 3.1). The technicians also report that a moderate amount of influence can be exerted on colleagues if necessary (scores around 3.3). The implementation of ProMES caused no significant changes in the perceived dependence and influence on colleagues.

Technicians report a limited dependence on their supervisor (scores around 2.3). If necessary, they can somewhat influence the supervisor (scores around 3.1). Working with ProMES did not change the technicians' perception of dependence and influence on the supervisor.

Relationship with supervisor. In general, the technicians of all three conditions perceive their ongoing interpersonal relationship with their supervisor as positive. So, in general, the effectiveness feedback meetings is not threatened by interpersonal problems between technicians and supervisors.

6.2.5. Summary

Both in the participation and in the transportation condition, the implementation of ProMES resulted in a large increase in the perceived amount and utility of feedback received. One of the few differences between the participation and control condition pertained to the perceived feedback utility, which was significantly higher in the participation condition and suggest that understanding and acceptance of feedback may be higher if the system is designed participatively. Mixed results were found regarding the goal criteria. Neither in the participation, nor in the transportation groups, did the implementation of ProMES result in more specific goals. However, goal difficulty increased in both conditions.

These results suggest that only partial fulfillment of the general design criteria was achieved. It seems that the productivity increases that were obtained have been caused by a combination of specific feedback and spontaneous goal setting. Since spontaneous goal setting does not guarantee that the goals set are specific and difficult, improvements in this area seem possible. However, the limited perceived controllability of results suggest that there are some barriers to effective goal setting.

6.3. ATTAINMENT OF PROMES DESIGN CRITERIA

Based on the literature on ProMES, a large number of design criteria was specified for the design process, the design, and the implementation of the design (Section 2.4.3). It would go too far to discuss each individual criterion at great length. Therefore, only the most salient criteria will be discussed.

6.3.1. Design process

Three design-process characteristics are considered essential for producing a high-quality system which is accepted by the personnel and management: participative design, review and approval by management, and discussion until consensus during design and review meetings. Directly after the end of the design process in Enschede and Utrecht (the 'participative' condition), the extent to which these characteristics were present was assessed by means of a questionnaire among the technician in these regions (response: 60 percent).

Participative design

Both in Enschede and Utrecht, the technicians were satisfied with the amount of influence they had had on the contents of the system (see Table 6.5). Their opinion had been taken into account, and the ideas of management did not dominate the 'prototype' system. According to the technicians, the facilitators had not tried to exert undue influence on the context of the system. The fact that the ProMES design process represented a departure from the rather autocratic style of decision-making management had used in the past, may have partly caused this positive result. Furthermore, this result may apply to the first two stages in the design process (establishing products and indicators), rather than to the subsequent stages (establishing contingencies and designing the feedback report). In latter stages, the input from the design regions was restricted, because situational characteristics–such as the complexity of the work flow and the interdependencies between the technicians–required a lot of preparatory work (e.g. operationalizing the indicators, collecting historical data for the contingencies) that could only be done efficiently by the facilitators. Although the technicians and supervisors did not regard this lack of involvement as a serious problem, it resulted in a lower level of understanding of the intricacies of the system than would have been desirable. In all, taking into account the complications caused by the situational constraints, a satisfactory amount of participation was probably achieved.

Table 6.5. Participative design (means and standard-deviations).

Item (1= strongly disagree - 5= strongly agree)	Enschede (n=14/74%)		Utrecht (n=10/48%)	
Influence on the contents of the system (0–100%)	69	(15)	70	(19)
Technicians' opinion sufficiently taken into account	4.3	(0.8)	4.4	(0.7)
Too much emphasis on opinion of management	2.4	(0.6)	2.0	(1.1)
Too much emphasis on opinion of facilitators	1.4	(0.5)	1.8	(1.3)
Controllability of results sufficiently taken into account	4.0	(0.8)	4.4	(0.5)

Review and approval by management

The prevailing opinion among the technicians present during the review and approval meetings was that management had adopted an positive and open-minded attitude. Rarely did management use its hierarchical position to have its preferred solution included in the system. Interestingly, the Enschede region did not get as good an insight into the priorities of management as the Utrecht region (Table 6.6). This was probably due to a lack of preparation by management for the first review and approval meetings (which took place in Enschede). This resulted in management taking up its position in an 'ad hoc' way (i.e. after a few minutes of deliberation). In the subsequent review and approval meetings (mostly in Utrecht), management knew better what to expect. In addition, the facilitators beforehand discussed the main 'bottleneck' issues with management.

Table 6.6. Review and approval meetings (means and standard-deviations).

Item (1= strongly disagree - 5= strongly agree)	Enschede (n=14/74%)		Utrecht (n=10/48%)	
Management's attitude during review and approval*	3.9	(0.7)	4.0	(0.7)
Better insight in management's priorities	2.5	(0.7)	3.5	(1.2)

* 1 =very unreasonable, 5= very reasonable

Discussion until consensus

Because of the size of the design team, some non-interactive decision-making techniques were used in the idea-generating stages of the design process. The results in the Table 6.7 indicate that there had been ample room for discussion; some of the discussions went on too long and could sooner have been satisfactorily concluded.

Together with the results from Tables 6.5 and 6.6, these results suggest that the overall decision-making strategy was one of joint decision-making (as defined in Section 3.5), in which the design teams could exert considerable influence.

Table 6.7. Discussion until consensus (means and standard-deviations).

Item	Enschede (n=14/74%)		Utrecht (n=10/48%)	
Length of discussions*	4.0	(0.7)	3.5	(1.4)
Role of facilitators (directiveness)**	3.3	(0.5)	3.3	(0.9)

* 1= too short, 3= right length, 5= too long,
** 1= too directive, 3= o.k., 5= not directive enough

6.3.2. Design

Each of the main system elements (products, indicators, contingencies, and feedback report) should meet a number of criteria. In addition, the process of measurement should be feasible.

Completeness of products and indicators

Based on the experiences with the provisional design, which contained 10 indicators and which was considered complete, three indicators were not included in the final system (percent return calls car stock, percent claims, and percent capacity used). The reasons for excluding these indicators were the inability to obtain consistently accurate data (percent claims, percent capacity used) and insufficient controllability (percent return calls car stock).

The Activity Based Costing analysis (used to aid in determining the relative importance of the indicators) revealed that the indicators to be excluded from the system were the least important ones from a cost point-of-view. Compared to the most important indicator (MCBC), percent capacity used, percent return calls car stock, and percent claims had an importance of 22, 16, and 1 percent, respectively. All three indicators were considered less important than the least important remaining indicators ('Accuracy of History Card' and 'Correctness of behavior' at 30 percent). These figures show that the completeness of the system was compromised only to a small degree by the removal of these indicators.

In the Nashuatec setting, another aspect of completeness has become relevant due to the complexity of the work flow as expressed in the different types of photocopiers (and fax machines, laser printers, etc.) serviced by the technicians. The ProMES system was designed for photocopiers only, since these reflected majority of the technicians' workload (hours spent per month). Figure 6.2 shows that technicians spent only nine percent of their time on non-copiers. Including these other machines in the system would have complicated the design, since some of the indicators established for photocopiers would not have been applicable to non-copiers.

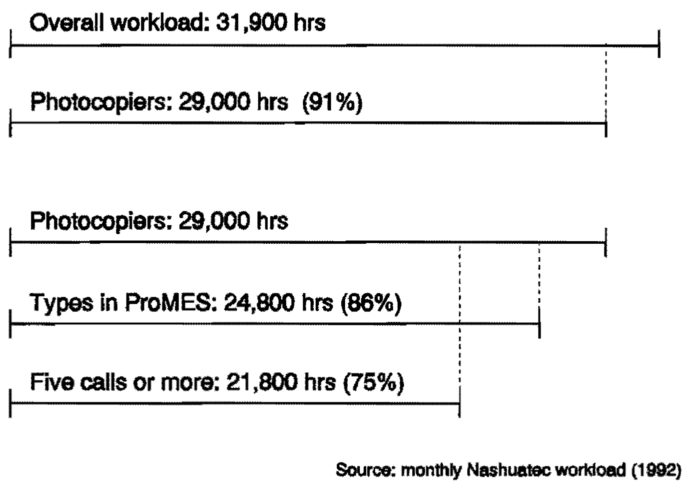


Figure 6.2. Portion of overall monthly workload covered in ProMES (1992).

Taking the workload on copiers as a starting-point, 14 percent of the workload was spent on types of copiers that could not be included in the system, because the limited amount of calls executed prevented determination of accurate contingencies[4]. An additional 11 percent of the workload would be 'lost', because a minimum monthly amount of five calls was required for inclusion in the individual feedback reports.

As shown in Table 6.8, ProMES covers most of the work of the technicians who originally worked at the Nashua company. This was not the case for the smaller group of technicians which had been added to Nashua's integration with another supplier of office machines (Gestetner). The main reasons for this were 1) these technicians worked on a large number of different types consisting of only a few machines each, which prevented the development of accurate contingencies, and 2) these technicians had a considerable workload on non-copiers (copy printers, stencil machines, velo-binds).

[4] Creating larger data sets suitable for establishing contingencies, by combining the data of comparable models, was considered. However, this turned out to be difficult to realize in the company's information system and was therefore rejected as an option.

Table 6.8. Part of total workload covered by ProMES (1992).

Part of work covered	All technicians	Original group of technicians (Nashua)	Additional technicians (Gestetner)
75–100 percent	33%	38%	2%
50–74 percent	33%	39%	2%
25–49 percent	20%	17%	39%
0–24 percent	14%	6%	57%

To summarize, the completeness of the system, defined as completeness of products and indicators, is satisfactory since only relatively unimportant indicators were discarded. However, for about a quarter of the technicians, ProMES is of limited utility because only a small amount of their work is covered by the system. This is a direct consequence of the 'complexity of work flow'-design problem and the organizational changes (integration of two field service departments) that occurred after the system had been completed.

Controllability of indicators

In general, indicators should be measures which are under the control of the unit. In the Nashuatec system, none of the individual indicators is under the complete control of an individual technician. Some examples: 'MCBC' is influenced by the handling of the machine and the copy volume of the customer; 'Percent repeat calls' is influenced by the malfunctions that are caused or might have been solved by the customer; 'Parts cost per call' and 'Labor time per call' are influenced by temporary technical problems on certain types of copiers, by prescribed replacement of spare parts and maintenance activities the maintenance procedure, and by the extent to which the previous technician followed the maintenance procedure. Whereas some of these influences will be about equal for most technicians if one considers a long period of time, other factors will not. For example, the random distribution among technicians of types of preventive maintenance–which differ in the amount of work that has to be done–may cause differences in certain months which will diminish over a long period of time (e.g. a year). On the other hand, the environment in which the machine is located, the handling of the machine by the customer, and the copy volume encountered by the technician may not be randomly distributed among technicians, since these depends on the types of companies in the technicians' normal working area.

Generating a reliable estimate of each indicator's controllability (the percentage of variance of effectiveness caused by knowledge, skills, and motivation of technicians as opposed to external factors (i.e. factors outside the control of the technician)) is not a simple task. To obtain a rough estimate of differences in the controllability of the four individual indicators, the between-subjects standard deviation was compared with the within-subjects standard deviation, using post-redesign effectiveness data (June 1991 through March 1992). If one assumes that variance between technicians is caused by differences in competency and/or motivation of the technicians (and ***not*** by external factors) and that variance within technicians is caused by external factors that vary on a monthly basis and have the same long-term effects for all technicians, the following rough index of controllability can be used:

$$\text{'controllability'} = \frac{\sigma_{between}}{\sigma_{within}} \qquad \begin{matrix}(n_{between} = 197 \text{ technicians}) \\ (n_{within} = 6\text{-}10 \text{ months})\end{matrix}$$

From the results in Table 6.9, one could conclude that 'Labor/time per call' is more 'controllable' than the other three indicators. The high between-subjects σ reflects large differences between technicians; the low between-subjects σ indicates that these differences are stable (e.g. 'fast' workers and 'slow' workers). One could also conclude that the 'Quality'-indicators are less controllable than the 'Cost'-indicators. This would be logical, since a number of uncontrollable events can occur between input (parts and labor time expenditure) and output (MCBC and repeat calls realized).

Table 6.9. A rough index of controllability.

Indicator (effectiveness range)	between subjects stand.dev. as percentage of range	within subjects stand.dev. as percentage of range	*sd*(between) / *sd*(within)
MCBC (200)	30 (15%)	25 (13%)	1.2
Percent repeat (140)	26 (19%)	22 (16%)	1.2
Parts cost/call (100)	17 (17%)	13 (13%)	1.3
Labor time/call (170)	33 (19%)	19 (11%)	1.7

However, the above line of reasoning might be incorrect if the between-subjects σ is in part caused by external factors that do not have the same long-term effect for all technicians (high/low copy volume? (un)favorable group of clients?).

During the design process, a minimum controllability requirement had been used ('a technician should control at least 50 percent of the outcome variability of an indicator used in ProMES') and subjective controllability estimates had been made for the indicators under consideration. The mean subjective controllability estimates for the four indicators discussed here were close together: MCBC: 72%; percent repeat calls: 64%; parts cost/call: 61%; labor time/call: 74%. Although these subjective estimates cannot be directly compared with the controllability index in Table 6.9, one might conclude that the MCBC-indicator has turned out less controllable than predicted, whereas labor time/call has proved more controllable than predicted.

Independence of indicators

Interdependence of indicators is not necessarily a problem. Indicators can be considered a chain of measures influencing a final measure (Pritchard, 1990 p.146). For example, a ProMES system for sales employees should not only look at a final measures such as 'number of products sold', but also at number of presentations made, number of prospects, amount of telephone work (cold calls), which all contribute to selling products. It would therefore be appropriate to measure at multiple points along the causal chain. In that case, a potential problem one should be aware of is that the function being measures potentially could be overweighted in the final measurement system. This can be prevented by first assessing the relative importance of the *products*, and using the outcome as a constraint in the process of scaling the indicators. Only if correlated indicators are really measures of the same thing, should one indicator be used.

During the design of the ProMES system at Nashuatec, 'MCBC' was considered the best overall indicator for the 'Quality'-product. Using 'MCBC' as the only quality-indicator was considered insufficient, because the time lag associated with this indicator meant limited utility for short-term performance management. As a consequence, a second indicator ('Percent repeat calls') was included as a short-term indicator for 'Quality', even though some overlap was thought to exist between both indicators (a high percentage of malfunctions within five working-days will involve a relatively low number of copies between two consecutive malfunctions). The overlap between both indicators was not explicitly taken into account in the financial analysis which was a partial basis for determining the relative importance of the indicators (Section 4.7.6).

Table 6.10 shows the correlations between the four type-dependent indicators for the post-redesign feedback period (June 1991–March 1992). As expected, there is a positive correlation between 'MCBC' and 'Percent repeat' (r=.446). There is also a positive correlation between the cost-indicators (r=.332). Latter correlation may indicate that a trade-off between labor time and parts replacement (e.g. increasing the lifetime of spare parts by cleaning them) has a smaller effect than increased (decreased) time consumption due to replacement (no replacement) of spare parts. The trade-off between quality and cost is reflected in negative correlations between the quality- and cost-indicators. This pattern of correlations is suggests that neither 'Quality' nor 'Cost' is overweighted in the system.

Table 6.10. Correlations between individual indicators.

Period: 10/90–5/91 (181 technicians)			MCBC	repeat	parts	labor
Quality	R =	MCBC		.446	−.288	−.299
		Repeat calls			−.225	−.327
Cost	−.406	Parts cost/call				.332
		Labor time/call				
Correlations significant at .001 level, except repeat calls–parts cost/call (.01).						

In case one should decide to account for differences in copy volume between machines by adding an indicator such as 'Time between calls' to the 'MCBC'-indicator already used, it would probably be appropriate to maintain the overall weight by distributing the original weight of MCBC among the two indicators, since a high correlation between these indicators is to be expected.

Contingencies

Accurate minima, zero points, maxima. A main weakness of the system is the limited utility of the contingencies to individual technicians. In order to compare technician's performance on different types of copiers, contingencies were developed for each type of copier included in the system. Historical data was used in order to efficiently and accurately establish the maxima, zero points, and minima. Although the contingencies were an accurate reflection of the range of indicators scores realized by the group of technicians as a whole, the contingencies did not reflect the range of indicator scores of *individual* technicians. As a logical consequence, about half the technicians received primarily negative feedback, which is often rejected as inaccurate (it violates the 'positive feedback'-rule). The bilateral feedback meetings incorporated an attempt to 'solve' this problem by having technicians establish their own performance range and setting goals in accordance with their own possibilities. These attempts have not been completely adequate.

To summarize, while the contingencies reflected the only practicable solution to the 'complexity of work flow' and 'individual task' situational constraints, there were concomitant negative effects.

Accurate reflection of relative importance of indicators. The ProMES system should be a reflection of policy, which is inherently subjective. However, this does not mean that no attempts should be made to quantify this policy in some way. The design procedure at Nashuatec incorporated a qualitative analysis of cost-effects of improvements on the indicators as input for establishing the relative importance of the indicators (see Section 4.7.6). In our opinion, this is an improvement upon the traditional (subjective) estimation procedure, since the 'estimation error' is limited to non-quantifyable outcomes such as customer satisfaction. Periodically, one should assess whether the relative importance of the indicators still accurately reflects company policy.

Feedback report

In essence, all desired elements but one (priority data) are included in the feedback report. A list of products and indicators is provided, together with the level of each indicator, its associated effectiveness value and the overall effectiveness value for the period. Because of the type-dependence of several indicators, some calculations are needed to determine the effectiveness value of these indicators.

Historical data was provided in the form of a six-month moving average. Although no direct comparison between monthly effectiveness and the moving average is given, this can easily be determined. Historical data over a twelve month period is given in graphical form. No percentage of maximum is calculated, because all the maximum (minimum) effectiveness value is +305 (−305) for all technicians, and no other units outside the field service department are using a ProMES system.

The final criterion involves priority data. Generally, ProMES offers a way to develop a set of priorities by presenting the effectiveness gains associated with improving each indicator value by one unit (Pritchard, 1990, p.100). By combining the priority data with information on the difficulty to make each change and the cost involved, strategies toward productivity improvement can be devised and implemented. This feature could not be included in the Nashuatec system, since there is no direct relationship between the overall indicator values and effectiveness scores, because of the type-dependency of the four individual indicators. Thus, simply calculating the effectiveness change per unit of indicator is not possible. An alternative would have been presenting the data for each type of copier, calculating the effectiveness value for each type, and setting priorities on the basis of an 'effectiveness-ranking' of the types. This option was rejected by management, because it would entail a substantial increase of the size of the feedback report and of the amount of programming and maintenance activities required.

Process of measurement

Supplying technicians and supervisors with the monthly feedback reports required regular assistance from the facilitator. This involved:
- collecting checklist data sent in by the supervisors, calculating indicator scores, and entering these scores into the ProMES program;
- executing the ProMES program for each region;
- checking the reports for completeness and accuracy.

The company's mail-room employees distributed the reports to the individual technicians involved (and a complete set of reports to the region supervisors involved). The regions supervisors were responsible for preparing the graphical feedback (on a monthly or bi-monthly basis).

In addition to these monthly tasks, the facilitator performed a bi-annual check of the accuracy of the contingencies. In case external factors–uncontrollable for the technicians–caused changes in effectiveness, the contingencies were adjusted. The main reasons for adjusting contingencies were changes in copy volume and higher maintenance requirements when machines have been used for a longer period of time. Finally, the facilitator periodically reviewed the total population of copiers to see which new types of copiers could be included in the system and which types should be deleted from the system.

The above activities were required to supply accurate monthly feedback in an unchanging environment. The complexity of the system required that a person with detailed knowledge of the system perform these activities. Therefore, the facilitator carried out most of these activities (which ultimately amounted to approximately 400 hours per year by the time all regions were using the system). Not included in this overview are activities consequent on changes in management policy, changes in responsibilities of service technicians, and use of the ProMES system for performance appraisal and incentive purposes (these activities also came to about 400 hours per year of facilitator-time).

In view of the fact that the ProMES system at Nashuatec provides monthly feedback to more than 200 employees, the resources needed and cost involved to do so are relatively small. An important advantage in this respect is the advanced information system Nashuatec uses, which enables 'automatic' data collection for the individual indicators by using the data technicians enter into the system by phone after each call ('voice data').

6.3.3. The implementation process: bilateral feedback meetings

The "Feedback Meeting Questionnaire" (FMQ)

To assess the effectiveness of the bilateral feedback meetings, a feedback meeting questionnaire was developed. The main focus of this questionnaire is the use of the key learning points (representing effective supervisor behavior) of the training program by the supervisors during the feedback meetings. The questionnaire contains five sections:

1. general information (e.g. number and length of meetings);
2. the introductory meeting (key learning points phase 1);
3. discussion of the feedback reports (key learning points phase 2);
4. setting quantitative goals (key learning points phase 3);
5. general characteristics (general key learning points).

Ten months after the start of the bilateral feedback sessions, the FMQ was distributed among the 163 technicians then involved with ProMES. The response was 82 percent (134 technicians). Data on the reliability of some scales used is presented in Appendix D. In the transportation condition, the questionnaire was filled in by the technicians at the beginning of the evaluation meeting. This ensured that their responses were not influenced by the discussions that followed.

Since ProMES is less relevant for technicians who mainly work on machines not covered in the ProMES measurement system, this group of technicians (31 respondents) was excluded from the analysis of the FMQ-results. The 103 remaining technicians all receive feedback on a substantial portion of their work (these technicians were all included in the analyses of productivity-effects in the previous chapter).

General information

The number of meetings and the progress made during those meetings has been somewhat disappointing (see Figure 6.3 and Table 6.11). Although almost every technician went through an introductory meeting, actual discussion of the feedback reports was done in only little over half of the cases. Almost no attempts were made to set quantitative goals. The technicians indicate that they would have liked more feedback meetings (extrapolation of the results yields a perceived optimum frequency of four meetings per year). The duration of the meetings (approx. 45 minutes) was considered satisfactory. As mentioned in Section 5.2.5, the supervisors were required to invest a lot of effort in working toward the service department's attainment of the ISO-9001 certificate during the first months of the period discussed here. This may explain the relatively small number of feedback meetings carried out.

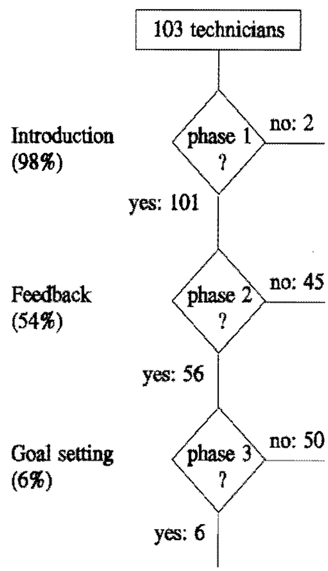


Figure 6.3. Progression of the bilateral feedback meetings.

Table 6.11. Feedback meeting questionnaire: general information.

General information	participation (n=29)		transportation (n=74)		partic vs transp
Number of meetings (in 9 months)	2.96[5]	(1.62)	1.92	(1.02)	P>T: .00
Frequency (1:too low - 5:too high)	2.07	(.77)	2.15	(0.84)	ns
Time spent per meeting	38 min	(23m)	46 min	(23m)	ns
Duration (1: too short - 5: too long)	2.82	(.39)	2.73	(.59)	ns
Progression of meetings					
started phase 1	28	(96%)	73	(99%)	
started phase 2	16	(55%)	40	(54%)	
started phase 3	1	(3%)	5	(7%)	

Numbers in brackets are standard-deviations or percentages / t-tests are two-tailed

[5] This may be an overestimate of the actual number of bilateral feedback meetings, since it probably includes some informal meetings in the participation condition before the start of the bilateral feedback phase.

Introduction (phase 1)

The first phase of the bilateral feedback procedure consists of one or more meetings in which the supervisor explains the objectives and contents of the program and tries to achieve a reasonable degree of acceptance on the part of the technician.

As shown in Table 6.12, the supervisors in both conditions implemented the key learning points for an effective introduction too a reasonable degree[6]. The supervisors in the transportation condition adhered closer to the key learning points than their colleagues in the participation condition. Presumably, latter supervisors assumed that their technicians–having participated in the design of the system–would be familiar with the objectives and contents of the program. The results revealed a difference of opinion on whether the supervisor presented the use of ProMES as a voluntary or an obligatory process. In the participation condition, the use of ProMES was predominantly considered mandatory, which is understandable since these technicians had been receiving feedback for 1½ years and ProMES information had been used twice in their performance appraisal. In the transportation condition, working with ProMES was predominantly presented as a more voluntary activity. However, large standard-deviations within both conditions (and within some regions) indicate that either the supervisors were not entirely clear regarding this issue or that they changed their position depending on the technician involved.

Table 6.12. Feedback meeting questionnaire: introduction (phase 1).

Introduction (phase 1)	KLP	participation (n=29)		transportation (n=74)		partic vs transp
Quality of introduction (5 items)	Phase 1	3.45	(.80)	3.81	(.77)	T>P: .04
Explains necessity	1b	3.21	(1.16)	3.59	(1.01)	
Explains objectives/contents	1c/d	3.21	(1.23)	3.67	(.99)	
Request technician's opinion	2	4.11	(1.07)	4.26	(1.00)	
Listens to technician's concerns	3	3.78	(.85)	4.18	(1.00)	
Achieves mutual agreement	5–8	3.04	(1.17)	3.33	(1.18)	
Technician may decide on working with ProMES		2.43	(1.53)	3.86	(1.53)	T>P: .00

KLP = Key Learning Points; numbers in brackets are standard-deviations / t-tests are two-tailed

Feedback (phase 2)

In the second phase of the bilateral feedback process, the feedback reports are discussed. After misunderstandings regarding the contents or accuracy of the feedback reports have been clarified, causes for high or low effectiveness scores are identified.

[6] Unless noted otherwise, the scores represent responses on a five-point Likert scales, such as those shown in Section 6.2.3.

Possible solutions are generated and arranged according to perceived contribution to productivity improvement. Agreement is reached on implementation of the best solution(s).

As can be seen in Table 6.13, the supervisors followed the key learning points for the feedback phase to a moderate degree (most scores between 3.2 and 3.7). There were no significant differences between the supervisors from the participation and transportation regions. On average, the technicians could not recall more than one agreement made during the feedback sessions. Closer analysis of the technicians' responses revealed that only about 60 percent of these agreement could be considered specific (e.g. "spend more labor time on type 4100 in order to improve your MCBC-score"; "make an inventory of the causes for your repeat calls"). About 40 percent of the agreements could be considered general (e.g. "keep up the good work"; "give more attention to this type of copier"). Even though few specific agreements have been made, the technicians report that it is moderately clear what should be done to fulfil the agreements.

Table 6.13. Feedback meeting questionnaire: feedback (phase 2).

Feedback (phase 2)	KLP	participation (n=16)		transport. (n=40)		part vs trans
Quality of explanation of feedback report	2	3.54	(.82)	3.80	(.76)	ns
Search for causes for low/high scores	3	3.56	(1.32)	3.60	(1.11)	ns
Search for solutions for causes (low sc.)	4a	3.56	(1.37)	3.63	(1.31)	ns
Attention to contribution of technician	4b	3.36	(.93)	3.54	(.89)	ns
Attention to contribution of others	4c	2.35	(1.00)	2.46	(1.12)	ns
Best solution chosen	6	3.21	(1.12)	3.26	(1.08)	ns
Technician's opinion taken into account	7	3.31	(1.14)	3.51	(.89)	ns
Number of agreements (0-3)		1.06	(1.18)	1.15	(1.04)	ns
Number of specific agreements (0-3)		0.59	(1.02)	0.61	(.95)	ns
Clarity of 'goal attainment' strategy		3.94	(1.18)	3.35	(1.21)	ns (.11)
Use of contingencies		2.31	(1.01)	2.55	(1.22)	ns
Use of graphical feedback		3.00	(1.56)	3.51	(1.36)	ns
Satisfaction with 'before' performance		3.00	(1.15)	2.58	(.98)	ns
Satisfaction with 'after' performance		3.50	(0.89)	2.89	(1.07)	p>t: .05

Numbers in brackets are standard-deviations / t-tests are two-tailed

The contingencies have not been used often during the bilateral feedback meetings. Generally, the contingencies aid in determining priorities through comparison of the effectiveness gain of equal increases (e.g. one unit on the horizontal axis) in the indicators scores. However, the complexity of the Nashuatec system–four contingencies for each type of copier–diminished the usefulness of the contingencies for setting priorities. Rather, simplified indices of relative importance were used (MCBC = 100, percent repeat calls = 70, etc.).

The graphical feedback was used more often, especially in the transportation condition. The graphs were considered useful for obtaining a quick overview of the effectiveness levels on the four type dependent indicators and the overall effectiveness score and for spotting trends in these scores. Even for technicians whose monthly effectiveness values fluctuated strongly due to a limited amount of work covered in ProMES, the graphs provided a stable estimate of their level of effectiveness. Some supervisors put more effort in producing these graphs than others; this is reflected by large standard-deviations.

At the beginning of the bilateral feedback meetings, the technicians were slightly dissatisfied with the ProMES effectiveness scores they attained (participation: 3.00; transportation: 2.58, as shown in Table 6.13). Eight months later, the mean satisfaction scores increased slightly (participation: +0.50; transportation: +0.31). These scores correspond with the slightly increased effectiveness levels in both conditions and the effectiveness increases that occurred due to the bilateral feedback meetings.

Formal Goal Setting (phase 3)

The majority of feedback meetings did not get past the second phase of discussing the feedback reports. Only in six cases (6 percent) was some form of formal goal setting attempted. This confirms the results from the Goal setting and Feedback Questionnaire (GFQ): the bilateral feedback meetings did not result in the setting of more specific goals. There is anecdotal evidence regarding two issues: first, both the technicians and the supervisors indicated that setting quantitative goals would be difficult unless the specific causes for certain scores on indicators and the impact of the removal of these causes would be known. This would require additional information (see also the sixth topic of the evaluation meetings in the transportation condition in Section 5.2.6). Secondly, most quantitative goals that were mentioned involved the zero point as a minimum level of acceptable performance. Technicians who do not attain the zero point are significantly less satisfied than their colleagues who have a mean score above zero ('after' scores of 2.39 and 3.38, respectively).

General Key Learning Points

In general, the supervisors have succeeded in maintaining a positive and constructive atmosphere during the feedback meetings by adopting a supportive attitude (Table 6.14). The feedback meetings turned out to be a relatively pleasant experience.

Table 6.14. General Key Learning Points.

General Key Learning Points	KLP	participation (n=28)		transportation (n=73)		diff.
Supportiveness of supervisor		3.94	(1.18)	3.93	(.83)	ns
"Focus on issues, not on persons"	1	3.94	(1.00)	4.13	(.99)	ns
"Stress the positive"	2	3.06	(.77)	3.20	(1.18)	ns
Pleasantness of meetings		3.75	(.65)	3.63	(.80)	ns

Key Learning Points are for 'Maintaining a constructive atmosphere', Table 5.7.
Numbers in brackets are standard-deviations / t-tests are two-tailed.

Attitude data

Overall, the technicians who participated in the design of the system expressed a moderately positive attitude toward the program, whereas the technician in the transportation regions expressed a predominantly negative attitude. As shown in Table 6.15, both the accuracy of the system and its usefulness for performance improvement were judged significantly more positive in the participation condition than in the transportation condition. Regarding the usefulness of the feedback meetings, the differences between the conditions were in the same direction, although somewhat smaller. The relatively high 'usefulness' (unspecified) of the bilateral feedback meetings may reflect its contribution to non-ProMES related topics, such as bringing day-to-day problems to the attention of the supervisor, and exchanging information on the ISO-project.

Table 6.15. Attitude data.

Attitudes towards ProMES/feedback meetings	participation (n=28)		transportation (n=73)		part vs trans
Accurate monthly reflection of performance*	3.56	(.72)	2.23	(1.10)	p>t: .00
Useful tool for maintaining/improving performance	3.37	(1.14)	2.64	(1.23)	p>t: .01
Useful way of using ProMES for performance improvement	3.37	(.88)	2.80	(1.14)	p>t: .02
Usefulness	3.78	(.75)	3.41	(.98)	ns (.08)

* = n=16 (participation), n=40 (transportation)

6.3.4. An illustrative example

To illustrate the importance of effective supervisor behavior during the bilateral feedback meetings, we will compare the two regions from the transportation condition that completed the second phase, 'Haarlem' and 'Maastricht'. Both regions started at an equally low level of effectiveness (−9), which enabled a direct comparison of effectiveness changes. The FMQ-responses indicate that the Haarlem supervisor acted more in accordance with the training's key learning points than the Maastricht supervisor

(Table 6.16). Especially during the feedback phase, the Haarlem supervisor was most effective in following the key learning points, using the material, and achieving clear agreements. This is also reflected in a sample of the agreements made (based on technicians' responses):

Haarlem: "Use more labor time to improve scores on MCBC (model 4100)"
"Adhere more closely to prescribed work routing and keep your spare parts box up-to-data in order to use your labor time more efficiently"
"Check the History Cards for causes of repeat calls"

Maastricht: "No mutual agreements were made, although my supervisor imposed some agreements on me"
"Pay more attention to certain types of copiers"

Table 6.16. Comparison of 'Haarlem' and 'Maastricht': feedback meetings.

	Haarlem (n=12)	Maastricht (n=13)	Haarlem vs Maastricht
General information			
Number of meetings	2.23	1.43	H>M:.02
Time spent per meeting	75 min	43 min	H>M:.00
Introduction (phase 1)			
Quality of introduction	4.37	3.96	n.s.
Technician may decide on working with ProMES	4.54	4.25	n.s.
Feedback (phase 2)			
Quality of feedback (KLPs followed)	3.67	3.08	H>M:.05
Use of contingencies	3.17	1.77	H>M:.01
Use of graphical feedback	4.25	2.50	H>M:.01
Number of agreements (0-3)	1.92	0.62	H>M:.00
Number of specific agreements (0-3)	1.58	0.31	H>M:.00
Clarity of 'goal attainment' strategy	4.17	2.58	H>M:.00
General Key Learning Points			
Supportiveness of supervisor	4.25	3.92	n.s.

Overall, the Haarlem technicians expressed a moderately positive attitude toward the program (comparable to the response in the participation condition); the Maastricht region, on the other hand, judged the program negatively (more so than the average response in the transportation condition) (Table 6.17). Although it might go too far to attribute this difference entirely to the bilateral feedback meetings, the evidence does point in this direction. There is also some anecdotal evidence that the supervisor's attitude toward ProMES may play a role in this respect. The supervisor of the 'Haarlem' unit expressed enthusiasm regarding the ProMES approach. During the training sessions, he also showed a relatively high level of proficiency at handling the bilateral feedback sessions. During the actual feedback sessions, he stimulated technicians to come up with their own action plans. The supervisor of Maastricht, on the other hand, expressed some doubts regarding the accuracy of the system and its practical implications (e.g. use of a computer program to generate graphical feedback), and achieved a lower level of proficiency than his colleague from Haarlem.

Table 6.17. Comparison of 'Haarlem' and 'Maastricht': attitudes toward the program.

Attitudes toward ProMES/feedback meetings	Haarlem (n=12)	Maastricht (n=13)	Haarlem vs Maastricht
ProMES accurately reflects monthly performance	3.00	1.46	H>M:.00
ProMES is a useful tool for maintaining/improving performance	3.38	2.21	H>M:.00
ProMES meetings are a useful way of using ProMES for performance improvement	3.53	2.50	H>M:.04
Usefulness	3.58	3.36	n.s.
Productivity change	+26	+13	n.s. (.31)

Finally, the productivity change (graphically represented in Figure 6.4), although not statistically significant, tends to confirm the above conclusion.

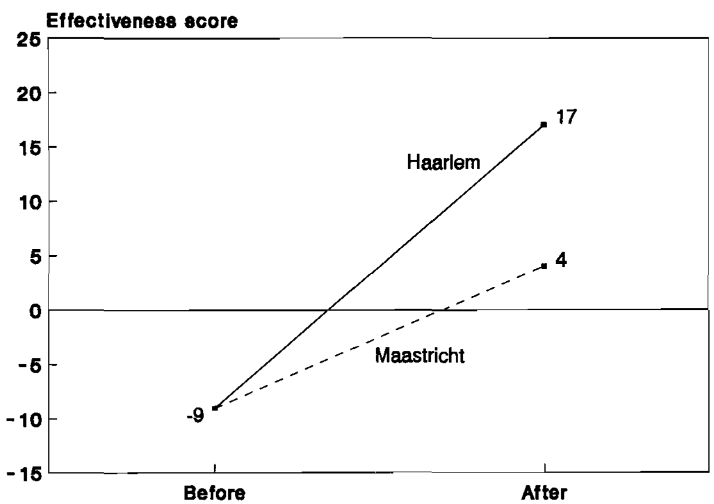


Figure 6.4. Haarlem and Maastricht: productivity change.

6.3.5. Implementation process: flexibility of the system

The ProMES system has proved to be flexible: several changes have been successfully incorporated into the system.

The first example refers to the type-dependent indicators. In november 1990, the first month of feedback, the system included contingencies for 25 different types of copiers. In the four year period that followed, 11 new types of copiers were added to the system and four types were removed. Only those types of copiers for which accurate contingencies could be established were added to the system. Two criteria were used for determining this. The first criterion refers to the number of calls that are performed on the type in question; if this exceeds 100 calls per month during a six month period, the type can be added. The second criterion referred the to the 'Mean Copies Between Calls'-indicator. For this indicator, a time-lag exists before indicator values can be calculated (see Section 4.4.2). Consequently, only less successful repairs –resulting in new calls during the next few months–will determine the mean MCBC-value which, therefore, is unrealistically low and based on a small number of calls. Only when the number of calls used for determining the MCBC-scores in a certain month approaches the number of calls performed in that month, will the MCBC-score be a realistic and can an accurate contingency be established. Depending on the malfunction-frequency, this takes between eight and twelve months. When a type of copier approaches the end of its economic lifetime, copiers of this type are gradually replaced by copiers of newer types. When this causes a workload of less than five calls for each technician working on that type, the type is removed from the system.

The problem of the moderating effect of the copy volume on the MCBC-scores (an important topic during the transportation phase of the project, Section 5.2.6), was partially solved by periodically adjusting the contingencies for the MCBC-indicator based on the observed change in copy volume. This is a second example of the flexibility of the system: Any external changes (outside the influence of the service technicians) which cause structural changes in indicator scores can be taken into account by adjusting the contingencies.

The high flexibility of the system is perhaps best illustrated by events that took place at the end of 1993. Gradual developments in the market had caused Nashuatec's management to shift its service policy from a mainly cost-oriented one to a more customer-oriented one. This necessitated several changes in the system, such as an increase of the relative importance of the 'Quality'-indicators compared to that of the 'Cost'-indicators (the quality-cost ratio increased from 170:135 to 165:80). Also, additional time was made available to all technicians for carrying out customer-related activities by adding 10 minutes to the range of 'Labor time/call'. The 'copy volume'-problem was addressed through the addition of a new indicator ('Mean Days Between Calls') to the system. Lastly, additional information was provided regarding causes for quality effectiveness. The first two of these changes could be implemented through adjustment of the contingencies. The latter two changes required some relatively small changes in the software program which generates the feedback reports.

6.4. SITUATIONAL CONSTRAINTS

In this section, we will summarize the design problems resulting from the main situational characteristics and the solutions that were implemented for these problems. The adequacy of these solutions in accomplishing an effective design will be addressed. The issue of consistency with other control systems–as brought out in the link between ProMES and the performance appraisal system–will be discussed separately in Section 6.5.

6.4.1. Individual task with interdependencies

An important design requirement was that the ProMES system should measure the productivity of individual technicians who, although working individually, are to some extent dependent on each other. This requirement posed several problems of a technical nature.

The participative design process was executed under the assumption that an individual system would require individual acceptance. Therefore, all technicians willing to participate were included in the design team. The large design teams that resulted necessitated the use of additional non-interactive decision-making techniques (Nominal Group Technique, Delphi technique, paired comparison technique). Although these techniques limited the amount of discussion, technicians indicated that their opinions had been taken into account sufficiently.

In disentangling the contribution of individual technicians to region productivity, some influence of colleagues (e.g. extent to which previous maintenance has been carried out) could not be excluded from the individual measurement system. In addition to external factors (customers, spare parts supply, etc.), these influences caused less than complete controllability of the indicators. Only those indicators were included in the system, for which an individual technician was assumed to control more than 50 percent of outcome variability.

Interdependencies were taken into account through the design of indicators stimulating cooperation between technicians (compliance with maintenance procedure, accuracy of History Card). Unfortunately, these indicators could only be measured by the supervisors through sampling. Because the small samples taken by the supervisors could not be used for feedback to individual technicians, the results were presented as region scores.

Partly because of the individual level of measurement, some indicators had to be excluded because their accuracy could not be guaranteed. Small errors that would not have had a noticeable effect on group scores became be unacceptable on an individual level of measurement (e.g. if a service planner enters a half day off for a certain technician too late (i.e. after the end-of-the-month deadline), this causes a small error in the region score of approximately 0.1 percent/2 effectiveness points, whereas the individual score of the technician in question would show a 2.0 percent/40 point-decrease).

When the final design was implemented by means of bilateral feedback meetings, individual indicators and individual productivity became the center of attention. Even though there were some comments suggesting that the system promoted competition among technicians which could harm the long-term quality of the machines, the effectiveness scores on the region indicators improved rather than decreased both for the participation and transportation condition, suggesting that long-term quality was not compromised.

In all, most design problems caused by the 'individual task/interdependence' constraint could be solved, albeit in a time-consuming design process. Two shortcomings of the system could not be prevented, namely limited controllability of some indicators and imperfect measurement of the indicators stimulating cooperation between technicians.

6.4.2. Complexity of work flow

The technicians carried out their work on a large number of different types of copiers, the technical characteristics of which influenced cost-investment (spare parts replacement, investment of labor time) and the quality (number of copies between calls, percentage repeat calls) that could be attained. The contingencies–a very useful feature of the ProMES method, because they enable comparison of scores on different indicators and the use of a single index of productivity–proved to be even more valuable in the Nashuatec context in resolving this 'complexity of work flow'-issue. By establishing contingencies per type of copier (based on historical data), indicator scores on different types of copiers could be compared directly. Also, productivity of technicians servicing an entirely different set of copiers could in principle be compared. Furthermore, including new copiers in the system and removing those no longer serviced was relatively easy.

The use of a standard procedure for establishing the contingencies (based on a historical distribution of indicator scores) prevented some of the problems associated with establishing contingencies: no adjustment of contingencies (due to incorrect subjective estimates) was necessary; there were almost no out-of-range indicator values; the minimum indicator scores which are generally difficult to conceptualize could be established with some certainty.

On the negative side, the feedback report turned out rather complex due to all kinds of calculations involved in aggregating indicator scores, providing moving averages, etc. Also, the large number of contingencies (four for each type of copier) decreased their usefulness for setting priorities. The contingencies as such were not suitable for individual goal setting, since they reflected the historical range of indicators scores of all technicians nation-wide, rather than the performance range of an individual technician. Therefore, an individual goal setting procedure was designed as a separate phase of the bilateral feedback meetings. Finally, only types of copiers with substantial workload could be included, whereas types of copiers with a small workload (no accurate contingencies) and non-photocopiers (indicators not applicable) had to be excluded from the system.

In all, the complexity of work flow problem could be solved satisfactory, although compromises had to be made regarding the accessibility and the completeness of the system.

6.4.3. Lack of horizontal communication

The lack of possibilities for horizontal communication, caused by the organization's geographical structure, necessitated some changes in both the design and the implementation process. The geographical dispersion of unit members prevented the use of a small design team consisting of unit representatives, because the necessary process of review and approval by the rest of the unit could not have been accomplished. This contributed to the design process discussed with the first context constraint (individual task/interdependencies) above, during which the facilitators fulfilled the additional role of messenger between the two design regions. Perhaps the most problematic consequence of the limited possibilities for communication was the limited involvement of the technicians in the operationalization of the indicators (at least the supervisors were involved in developing the checklists for the region indicators) which caused a temporary lack of insight into the system. During the feedback phase, the geographical dispersion of technicians presented the supervisor with practical difficulties in conducting the bilateral feedback meetings.

In all, the lack of horizontal communication possibilities mainly caused practical inconveniences.

6.4.4. Lack of trust

In general, some degree of mutual trust and respect between group members, supervisors, and management should be present, if the design and implementation of ProMES is to meet with any success. At the start of the design process at Nashuatec, this precondition was not fulfilled. One likely reason was the lack of vertical communication between management and design regions, due to which all kinds of prejudices arose and continued to persist. Another probable reason was the top-down culture within the service department, with managers making decisions pertaining to the service regions without consulting them or explaining the rationale behind these decisions. Furthermore, the service technicians may not have been able to identify with an organization whose head office they hardly visited.

In the participative design process, this mistrust was reflected by the written guarantee that had to be issued by management as a prerequisite for starting the design process in Enschede/Utrecht: a future ProMES system would not be used for performance appraisal purposes unless both the regions involved and the management would consider it sufficiently accurate to do so. Furthermore, the guarantee was given that both regions would have the opportunity to withdraw from the project at any stage if it were to turn out impossible to achieve an acceptable design. Notwithstanding these guarantees, feelings of mistrust toward management persisted during the design process. The review and approval meetings were somewhat useful in reducing these feelings of mistrust, because they provided the technicians with some insight in the priorities and motives of management. Management reported to be impressed by the results achieved by the design teams.

During the research project, the relations between management and service regions seem to have improved, perhaps due to intensified communication (e.g. field service managers attended region meetings more frequently, the ISO certification effort required a lot of additional communication) and a tendency for management to seek the opinion of the service regions in decisions they would have made unilaterally in the past.

The quality of the relationship between a technician and his supervisor may moderate the effects of the bilateral feedback meetings (for instance, one can hardly expect a constructive problem-solving approach to productivity improvement if both parties are at a 'state of war'). A large majority of the technicians reports at least a reasonably positive relationship with their supervisor. Therefore, this precondition for effective implementation of the system appears to be fulfilled.

To summarize, the apparent lack of trust in the intentions of management (especially regarding performance appraisal consequences) hampered the design process. The review and approval meetings may have lessened these feelings of mistrust. Improved relations between service regions and management and reasonably positive relations between individual technicians and supervisors have since created a better climate for working with ProMES.

6.4.5. Other situational constraints

Some other situational constraints discussed in Chapter 2, referring to the state of the organization and the attitude of the organization towards productivity and ProMES, were relevant in the Nashuatec context.

Stability of the organization's management. There were no personnel changes in middle and upper service management. There were, however, some changes in the service regions. During the research period, the size of the field service organization increased rapidly (number of technicians (regions): 1989: 220 (9/11), 1990: 230 (11), 1991: 290 (11/14), 1992: 290 (14/13), 1993: 270 (13)). Among the reasons for this increase were the growth of the market for office equipment and the additional group of technicians that entered the company in 1991. These changes involved re-allocation of technicians and supervisors to differing numbers of regions and the selection of new supervisors and technicians.

Apart from some practical problems (e.g. keeping track of who had been introduced to ProMES and who had not), the most problematic consequence of the integration was the sudden introduction of a large number of copiers that could not be included in ProMES, due to a different method of maintenance and an insufficient workload to establish accurate contingencies. The technicians who mainly worked on these machines did not receive feedback from ProMES until after they started working on types of copiers included in the ProMES system.

Management commitment to the ProMES program. Visible management commitment to the ProMES program is considered of critical importance to its success. Nashuatec's management had taken the position that the system should be able to 'pay it's way', i.e. function self-reliantly, without regular attention for management. This lack of commitment was one of the main perceived causes for the initial lack of productivity improvement in the participative condition. Although management supported the changes that were required to bring the system into agreement with the design criteria, ProMES was long seen as a peripheral activity with minimal involvement of management, rather than a central part of the service organization's operations. This became apparent, when it took several months for the service management to realize that the potential cost reduction that resulted from time gained through more efficient use of labor time–a substantial portion of the overall effects of ProMES on productivity–would equal zero actual gain if technicians were sent home before the official end of the working-day. Ultimately, the 'idle' time was used to enhance customer satisfaction (by carrying out preventive maintenance on copiers that had not been visited for at least a year).

Familiarity with measurement. In large part, the design requirements posed by the 'individual task/interdependencies-' and 'complexity of work-' situational constraints could be realized thanks to the sophisticated information system and data-processing expertise available at Nashuatec. If this had not been the case, it would not have been possible to provide detailed monthly feedback to more than 200 individual technicians while taking into account the characteristics of the different types of copiers. Some of the advantages of this situation were: availability of raw data for several indicators (which could be retrospectively converted into indicator data for the ProMES system), availability of data for the strategic cost-analysis, and availability of data for checks and balances (travel time, special time-uses) and external influences (copy volume). Except for the group indicators, no additional measurements had to be done, since all data required was entered into the system by the technicians themselves through a 'voice data' system.

Possibility of substantial productivity improvement. The possibility of substantial improvement through increased motivation will influence the willingness of management to support the program. In the Nashuatec situation, it was clear that 'technical' interventions which, among others, resulted in uniform maintenance procedures, a sophisticated information system, an inventive way of delivering spare parts, and effective planning procedures, had produced a high base-level of productivity. However, apart from some previous success in reducing the amount of repeat calls, no 'performance management'-interventions had been successful. Management did have some indications that there was room for improvement, mainly regarding efficient use of inputs (labor time, spare parts).

Based on the knowledge obtained in the participation and transportation of the system one can ask oneself what maximum productivity gain would be feasible. Unlike the ProMES system at the US Air Force, there are substantial trade-offs between the indicators (see the negative correlations between the 'quality' and 'cost' indicators in Table 6.10). These suggest that maximum productivity does not approach the theoretical maximum (the sum of all indicators maximum effectiveness values) of 305. Some interesting mean effectiveness scores realized in 1992 were: 9 (mean productivity nation-wide), 100 (the 'best' technician), –80 (the 'worst' technician), and 35 (the 'best' region). Keeping in mind the productivity mean productivity increases thus far in the participation condition (+25) and the transportation condition (+11). Based on these data, a rough estimate of maximum 'nation-wide' level of productivity could be somewhere between 30 and 50.

6.5. USE OF PROMES INFORMATION IN THE PERFORMANCE APPRAISAL

In Section 4.7.5, a brief description was given of the use of ProMES information in the performance appraisal process, as agreed upon in the participation condition. There were two main reasons for linking ProMES to the appraisal system. First, the performance appraisal system is an important control system, which–like ProMES–aims at stimulating employees to contribute to attaining organizational goals. Whereas ProMES takes a short-term perspective (through monthly feedback), the performance appraisal system takes a more long-term perspective (through an annual appraisal session). Both systems should be consistently linked to fulfill their functions effectively. The second reason was a practical one. The region supervisors, who would be using the ProMES reports of their technician in the context of the bilateral feedback meetings, would not be able to ignore this information, even if they would be required to do so. Therefore, both management and design teams agreed to develop a procedure which was to guarantee uniform and equitable use of ProMES information in the performance appraisal.

6.5.1. The Nashuatec performance appraisal system

In 1991, the personnel department introduced a new performance appraisal system. The system–although an improvement over the system used previously and representative of systems used in other large Dutch organizations–possessed several of the typically dysfunctional characteristics of performance appraisal systems (Janssen, Van Tuijl & Algera, 1987; Latham & Wexley, 1981; Van Tuijl, Janssen, Algera, 1987). These characteristics refer to the appraisal system itself, the appraisal process, and the consequences of the performance appraisal.

Characteristics of the system

- The performance dimensions were formulated in general non-function-specific terms. Therefore, they would be an insufficient basis for specific feedback or specific goal setting.
- The performance dimensions were mainly trait dimensions ('task conception', 'self-reliance', 'flexibility'), rather than specific behaviors or outcomes. Trait dimensions are ambiguous, because they do not tell the individual what to do to improve and, therefore, do not allow for specific feedback and goal setting.

- No explicit standards were stated for what would constitute performance which 'greatly exceeds demands', 'exceeds demands', 'meets demands', 'does not meet demands' or 'is far below demands' (which were the anchors used in the Nashuatec system).
- There was no importance weighting of the dimensions, which means that each appraiser could apply his/her own subjective weighting and which suggests that all dimensions apply to all functions.

Characteristics of the performance appraisal process
- No uniform, systematic, registration of performance data was used, which increases the risk that recent occurrences carry too high a weight in the overall appraisal.
- The appraisal frequency employed–once a year–is generally considered too low to have a substantial effect on work behavior.
- Several rating errors were made by the majority of appraisers (central tendency, halo effect, contrast effect, et cetera).
- On a more positive note, some attention had been paid to some of the skills the supervisors would need in the performance appraisal process by means of a 'situational leadership' and 'performance appraisal' training programs. The effects of these training programs may have been limited since the preferred 'problem-solving approach' to the performance appraisal process is difficult if its basis–a valid measurement system–is lacking.

Consequences of the performance appraisal. There was hardly any differentiation in rewards: the amount of bonus paid was the same for 70 percent of the technicians. This is typical of 'merit rating' systems (Thierry, 1987). Because of the often vague dimensions used, differences between employees are not easily defendable, which causes appraisers to differentiate only in extreme cases. This behavior is reinforced by the fixed available budget which often causes a forced normal distribution concentrating around the mean. The trade-off between the amounts of bonus for individual employees does not stimulate differentiation, since an increased bonus for one employee automatically causes a decrease in bonus for another employee (whereas former increase is easily defendable, latter decrease may meet with resistance if vague dimensions are used).

Improvements made

At the request of supervisors, service management and facilitators, three changes were incorporated into the appraisal system for the service technicians in order to solve some of the above problems.

First, importance weights were added to the appraisal dimensions. Dimensions that were considered important were given a high multiplication factor (e.g. task conception, quality of work), while less important dimensions were given a low multiplication factor (e.g. vertical communication). Also, two dimensions were given a zero multiplication factor, because they did not apply to the function of service technician (leadership, helicopter view). The weights were determined by the supervisors and management representatives from the service and personnel departments.

Secondly, a concrete description of the dimensions applied to the function of 'service technician' was added to the appraisal form in order to clarify the meaning of the dimensions.

The third change involved the bonus outcome of the appraisal process to which ProMES would be linked. Originally, five bonus-levels could be used (0, 2, 4, 6, and 8 percent of the gross annual salary, corresponding with an overall appraisal score of 1, 2, 3, 4, and 5 respectively). In practice, almost 70 percent of the technicians received a four percent bonus, although the technicians at the higher end of this category performed clearly better than their colleagues at the lower end. According to the supervisors, this was caused by the large difference between the 'meets demands' and 'exceeds demands', which were the respective definitions of a four percent bonus and a six percent bonus. Therefore, the number of bonus-levels was increased to nine (0, 1, 2, 3, ..., 8 percent). In particular, the availability of the three and five percent bonus outcomes would enable some differentiation in the bonus percentages paid. Performance differences between technicians who formerly all received a four percent bonus could now be shown by giving them three, four or five percent.

6.5.2. Incorporating ProMES information into the performance appraisal

The procedure for incorporating ProMES information into the performance appraisal, agreed upon by management and the two design teams, was the following. ProMES contributed to the distribution of the annual bonus (varying between zero and eight percent of a year's salary) and not to annual salary increases. According to MacLean (1990), this is a correct reflection of the distinction between accomplishments and competencies. Accomplishments are temporary and variable. Because of this volatility, it makes little sense to reward them on any other basis than once as they are achieved (e.g. through a one-time bonus). Salary, on the other hand, should be based on the employee's competencies, which are relatively stable and permanent/additive over time.

Seven dimensions of task accomplishment were perceived to be related to the ProMES outcome measures to some extent, whereas three others were not. Taking into account the weights added to the dimensions, the ProMES-related dimensions accounted for 80 percent of the overall appraisal. It was decided to determine one half of this percentage by the supervisor and the other half by information from the ProMES system. This means that, in all, 60 percent of this bonus would be determined by the supervisor's appraisal of the technicians performance, using the standard performance appraisal form. The remaining 40 percent would be determined by the ProMES information.

The decision was made to use the effectiveness scores on the four individual indicators as a basis for the ProMES part of the appraisal. One half of the forty percent would result from a comparison of the technician's average absolute effectiveness value in the appraisal period to that of all other technicians. The other half would be determined by the technician's relative effectiveness score (i.e. the difference between the period at hand and the period covered in the previous appraisal) compared to the relative score of all other technicians. As a result, technicians are rewarded for the absolute level of productivity they achieved in the appraisal period, as well as for the amount of productivity change compared to the previous period. Solely using the absolute score would make the amount of bonus highly dependent of the technician's skills, knowledge and experience. As a consequence, highly motivated but less experienced or less proficient technicians would not be able to obtain a medium of high amount of bonus. (Taking the point of view of MacLean (1990), differences in competency should mainly be reflected in the level of salary.) Solely using a relative score would be demotivating for technicians who have already achieved a high level of productivity and for whom further improvements beyond this ceiling would be difficult (most of these technicians would not get a medium of high bonus). Combining the absolute and relative scores diminishes the drawbacks of using either one alone.

A global outline of the use of ProMES in the performance appraisal is shown in Figure 6.5. Figure 6.6 shows the relationship between the appraisal dimensions and the ProMES-outcomes.

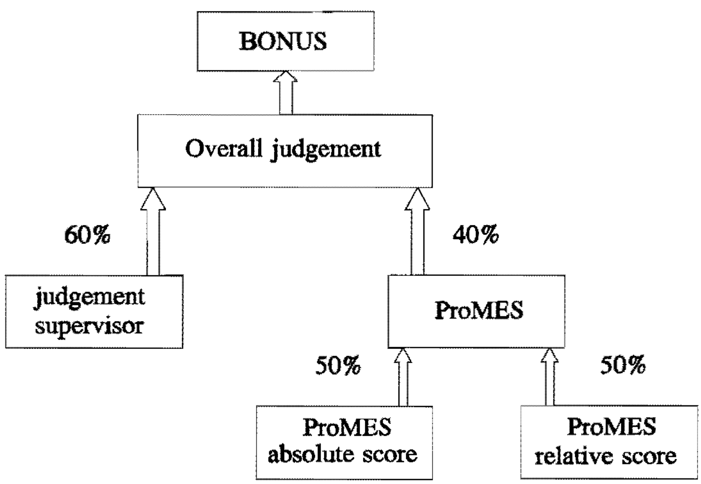


Figure 6.5. Contribution of ProMES to the annual bonus.

By agreement with the labor union, the mean amount of bonus in all departments was limited to five percent of the gross annual income of the employees in those departments. Therefore, the mean bonus paid to all service technicians should not exceed five percent, irrespective of the actual level of performance/productivity. ProMES could only contribute to a more equitable division of the fixed overall amount of bonus, and not to an increase of the overall amount. With this constraint in mind, both the absolute and the relative appraisal results of an individual technician were determined by

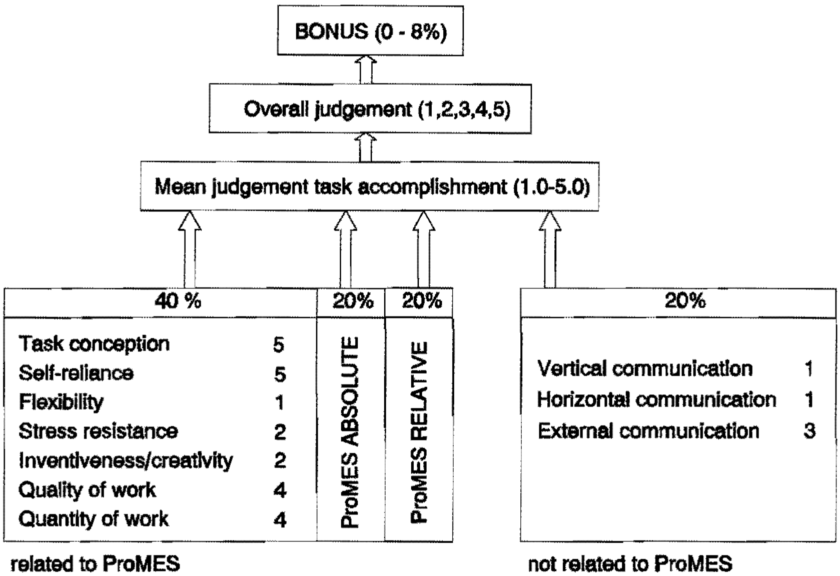


Figure 6.6. Relation of ProMES to performance appraisal dimensions.

his position in a distribution of the effectiveness scores of all relevant technicians[7] (see Figures 6.7 and 6.8). In both distributions, the mean score corresponded with a bonus percentage close to five percent. Each score in the range corresponded to a bonus percentage between zero and eight percent.

In October 1992, the above procedure was followed in the two participation regions. Except for a few minor points, this procedure was identical to the test-appraisal the Enschede and Utrecht regions had agreed upon one year earlier. Technicians in the six transportation regions could decide for themselves whether the ProMES information should contribute to the performance appraisal (only those technicians whose reports represented more than 25 percent of their total workload were eligible). 31 percent of these technicians agreed to the use of ProMES information in their appraisal, 69 percent preferred postponement). Table 6.18 shows how the different elements of the appraisal contribute to the overall bonus.

[7] ProMES did not contribute to the performance appraisal of those technicians of whom ProMES covered less than 25 percent of their work in the twelve-month period under consideration or in more than four individual months in this period.

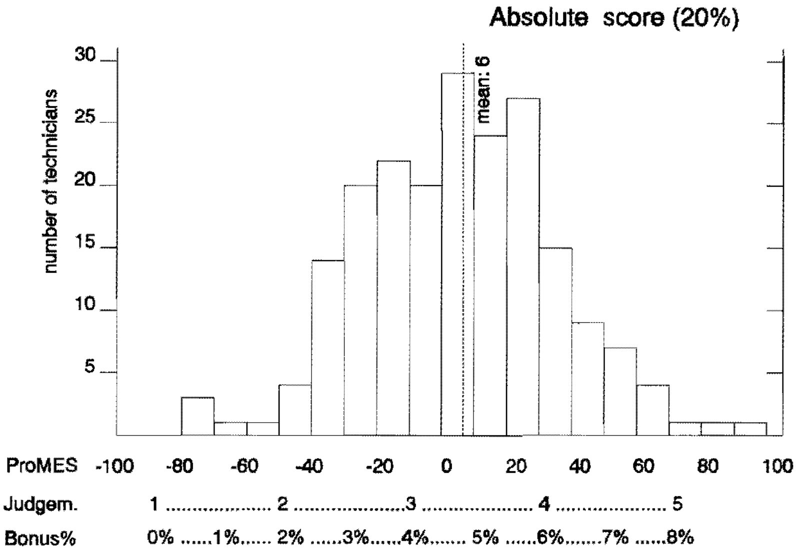


Figure 6.7. Conversion of absolute effectiveness scores.

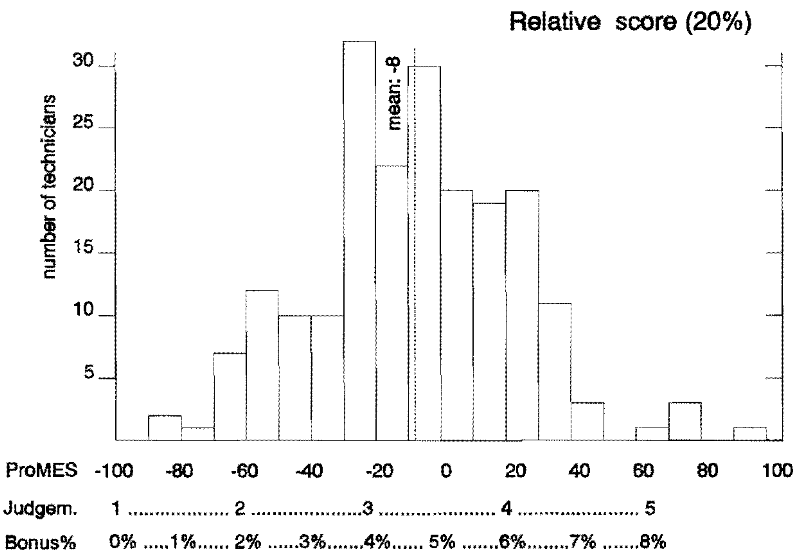


Figure 6.8. Conversion of relative effectiveness scores.

Table 6.18. Composition of the overall judgement and bonus (if ProMES is used): an example.

Composition of overall judgement/bonus (version: including ProMES information)		effectiveness score	appraisal score
ProMES- absolute (this period)		35	(4.13)
ProMES- relative (this period minus previous period)		10	(3.85)
ProMES - overall [(abs + rel)/2]	40 percent		3.99
Supervisor judgement - related dimensions	40 percent		3.74
Supervisor judgement - unrelated dimensions	20 percent		2.90
Overall judgement/bonus			3.67 (5.34% = 5 %)

6.5.3. Evaluation

As reported in Chapters 4 and 5, the use of ProMES in the performance appraisal had been the most controversial issue in the entire project.

For two reasons, the issue of performance appraisal had been relevant from the start of the project. The first reason was the secondary objective of the system as expressed by the management: 'if the ProMES system were to measure the performance of individual technicians in a valid way, it would be worthwhile to use the information generated by ProMES in the performance appraisal of the technicians'. The second reason was the desirability (from a performance management viewpoint) of measuring the performance of individual technicians, rather than just the performance of entire regions.

Basically, two viewpoints can be taken regarding the use of ProMES-information for performance appraisal purposes: the 'consistency between control systems view' and the 'incompatible tendencies view'.

Consistency between control systems

The 'consistency between control systems view' (already mentioned in Section 2.4.2 as an implementation criterion), basically argues that the more valued outcomes are associated with feedback, the higher the perceived importance of feedback will be, and the lower the likelihood that the feedback will be ignored. Consequently, inconsistency between feedback systems and reward systems can have dysfunctional effects. For example, if the organization wants to stimulate high performance on dimensions A, B, C, D, and E (perhaps through a ProMES system in which these result areas are the indicators in a monthly feedback report), the organization should reward the performance on all of these dimensions, and not on dimensions A, F, G, and H (for example). This inconsistency between the performance management system and the performance appraisal and rewards system will most likely cause employees to ignore feedback on dimensions B through E, which will cause suboptimal performance (if one assumes that the performance management system correctly reflects management policy with regard to the employees working with the system).

Incompatible tendencies

The other line of reasoning, the 'incompatible tendencies view', which ultimately discourages the use of ProMES information for performance appraisal purposes, is given by Van Tuijl (1990). Performance management systems (such as ProMES) and performance appraisal/reward systems pursue the same objective, namely stimulating employees to focus their attention on contributing the performance the organization asks for and on developing of more effective and efficient strategies for realizing this performance. There are, however, two differences between the two types of systems. Whereas a performance appraisal/rewards system typically operates on a long-term basis (e.g. annually), a performance management system typically takes a short-term (weekly, monthly) perspective. The second difference pertains to the operational use of performance-information. In case of performance management, performance information is used as (in itself) neutral feedback within a control loop. In the context of a performance appraisal, performance-information is used as a basis for determining the reward given by the organization to the employee in exchange for the contribution realized. In both systems, the performance-information has to conform to essentially the same requirements; these include responsibility, relevance, controllability, validity (completeness, accuracy, relative importance), cost-effectiveness, and understandability. For employees who want to regulate their own performance, some deviations from these criteria may be acceptable: a hundred percent controllability is not feasible, effects of external factors will average out in the long run, some subjectivity in measurement and evaluation cannot be prevented, etc. At the same time, there is a high need for valid performance-information, and consequently, little inclination to distort the information. Those who are appraised, on the other hand, may be inclined to demand that only those performance dimensions that are completely controllable by the individual are used in the appraisal, and put forward that one cannot just assume that the effects of external factors average out, etc. There may be a tendency to exert influence on the performance measurement to create a favorable starting-point in the performance appraisal. In other words, two incompatible tendencies result from the two types of systems:

* The practical aim within a performance management context is obtaining comprehensive and accurate performance-feedback on all areas of responsibility.
* Within the performance appraisal context, there is a need to positively influence the exchange ('quid pro quo')-relationship between the individual and the organization.

Only if sufficient guarantees are given for the validity of the performance-information ('checks and balances'), is joined use of both systems conceivable.

In case a performance system is designed participatively (e.g. ProMES), an implicit or explicit secondary objective of using the system for performance appraisal purposes may be a severe handicap. It may interfere with the open-minded generation of products and indicators (they may not be completely controllable, etc.) and cause an inclination to establish 'safe' contingencies. Reaching consensus, especially during review and approval meetings with management, may become difficult, since the 'quid pro quo'-relationship between the individual and the organization remains an issue.

A dilemma

Essentially, the combination of these views seems to present a no-win situation in the Nashuatec context. The situational characteristics, relevant for the performance appraisal issue, are the following:
* The technician's job is an individual one, which means that feedback on individual performance is indispensable for achieving performance improvements. The ProMES system therefore provided individual feedback in addition to group feedback.
* There is some interdependence between technicians (they may successively visit the same machine/customer). This meant, for instance, that controllability of indicators by an individual technician would be less than a hundred percent.
* There was some mistrust in the (design) regions regarding management's intentions, which made the issue of performance appraisal even more sensitive issue.
* Even if the management had not explicitly mentioned linking ProMES to performance appraisal of service technicians as a secondary objective, it would have become an issue anyway, because the supervisors would not be able to ignore the ProMES-information on the performance of individual technicians they would use during the feedback meetings. Likewise, some technicians would undoubtedly have asked for inclusion of their ProMES information in their performance appraisal.
* Because of the interdependence between technicians, some cooperation between technicians was required. Unfortunately, the extent of cooperative behavior could not be attributed to individual technicians by means of individual indicators.
* Although some 'checks and balances' were used to trace manipulation of the measurement information, they were by no means comprehensive.

Taken together, these characteristics seem conductive to the negative consequences predicted by the incompatible tendencies view. On the other hand, preventing these consequences from occurring would necessitate relinquishing individual performance measurement. Only the unavailability of individual performance information would prohibit use of the ProMES system for performance appraisal purposes; however, the lack of regular individual feedback would virtually preclude any performance improvements, rendering the ProMES system useless for achieving its main objective of productivity improvement.

Experiences in the participation and transportation condition

The experiences with the performance appraisal issue have not been the same in the participation and transportation conditions. In the participation condition, the design process was hampered by the performance appraisal issue (even though management had issued the written guarantee mentioned earlier). Ultimately, however, the participation regions agreed to the use of ProMES-information in the performance appraisal.

The reasons for this were a reasonable degree of confidence in the validity of the system, the realization that the ProMES-information would influence the supervisor's appraisal of the technicians anyway, and the realization that a uniform use of ProMES would increase the quality of information used for some dimensions of the appraisal. Although a direct causal relationship between the use of ProMES in the performance appraisal and the productivity increase in the participation condition cannot be proved (too many changes were made concurrently), there are indications that at least part of the increase was due to the agreed-upon use of ProMES in the performance appraisal. For instance, the 10-point productivity increase (Figure 4.7) in the three month-period after the agreement had been made (in the beginning of June 1991) may very well have been due to the expected use of ProMES in the performance appraisal, since it could not have been due to increased quality of the feedback (feedback was not resumed until September 1991).

In the transportation condition, the performance appraisal issue was the main stumbling-block for acceptance of the system. A recurrent remark in this respect was: 'ProMES is valid and useful for regulating our performance; it is not suitable for the performance appraisal'. The reasons for this were not entirely clear and seemed to be a combination of several factors. These included:

Limitations of the system:

* limited controllability of some of the indicators;
* incomplete measurement of a technician's performance (fax machines, laser printers, and some types of copiers were excluded from the system);
* individual circumstances not taken into account sufficiently (because 40 percent of the appraisal is predetermined by ProMES).

Negative consequences:

* fear of manipulation of the data by less scrupulous colleagues;
* fear of increased competition between technicians which negatively effects both the atmosphere in the region and the long-term condition of the copiers;
* increased perceived pressure to 'score' (especially in case of a 'negative' baseline effectiveness).

The first three factors are the least impressive ones, since it is highly doubtful whether the 'indicators' the supervisor uses in his subjective appraisal are anywhere near as controllable, complete or accurate as the ProMES indicators. The other three arguments referring to negative consequences seem to be the most important ones, since these consequences might not occur if the supervisor were the only source for the appraisal.

6.5.4. Further developments

Although this thesis covers the period from April 1989 (entry into the organization) through February 1993 (evaluation of the transportation process by the supervisors), in order to round off this discussion of the performance appraisal issue, we will mention one development that took place in December 1993 after the system had been introduced in the entire field service department.

Because the performance appraisal issue had remained a source of dissatisfaction with the ProMES system in the regions that had not participated in the design of the system, a project team consisting of representatives from all regions (one technician each per region and four supervisors) and the researcher tried to resolve this issue. This group unanimously suggested a less 'deterministic' application of ProMES in the performance appraisal by discontinuing the '40-60 procedure'. Instead, the following procedure was recommended:

1) The supervisor uses the objective ProMES-information to reach a judgement on the 'quality' and 'quantity' dimensions of the appraisal. In interpreting the ProMES information, the supervisor takes into account the specific circumstances of the technician and the portion of his work covered by ProMES.
2) The way the technician uses ProMES (e.g. effort put into bilateral feedback meetings, motivation to improve performance, and fulfillment of agreements) is incorporated in the 'task conception'-dimension.
3) Limiting conditions:
   - guidelines should be drawn up for uniform application of 1) and 2);
   - bilateral feedback meetings should take place at least once every three months.

This procedure for using ProMES in the performance appraisal would constitute a more logical conclusion of the bilateral feedback meetings during the past year. According to the project team, the limitations of the system would be less of a problem. Likewise, the risk of negative consequences would be reduced. Because the ProMES-information still contributes to several important dimensions in the performance appraisal system, there may still be 'enough' consistency between ProMES and the performance appraisal system to stimulate efforts aimed at improving productivity. Perhaps this compromise will provide a way out of the dilemma posed by the characteristics of the Nashuatec context.

6.6. USE OF PROMES IN THE SUPERVISORS' PAY-FOR-PERFORMANCE SYSTEM

6.6.1. Background

At the time preparations were made for the transportation of the ProMES system, a pay-for-performance ('management by results') system for the supervisors and middle managers in the service department had been in operation for about a year. Key result areas had been established, indicators had been formulated for each of the areas, and a three levels of goals, corresponding with three bonus-levels, had been set. For the supervisors, indicators included: percent return calls, percent repeat calls, hours per million copies, mean number of calls per day, number of after visit checks performed, and number of service calls carried out. These indicators determined 75 percent of a substantial annual bonus (the regular performance appraisal contributed 25 percent).

The experiences with the pay-for-performance system for the supervisors had not been satisfactory. The measurement system did not take into account differences in machine population between regions (e.g. some regions had an older machine population than others). Furthermore, unpredictable external developments had prohibited the setting of realistic performance goals on most indicators; as a consequence, some of the goals had been adjusted during the measurement period and even after the period had been completed. Management and supervisors were of the opinion that ProMES would provide more accurate information for some of the indicators in the pay-for-performance system. Likewise, using ProMES in the system would cause a higher consistency between the priorities in the performance management (and in some cases, appraisal) system of the service technicians and the priorities of the supervisors. In particular, the link between ProMES and the pay-for-performance system would stimulate the supervisors to aid their technicians in working with ProMES by giving sufficient attention to the bilateral feedback meetings.

The rationale behind this approach is confirmed by Pritchard (1990, p.126), who argues that using ProMES in the performance appraisal of the unit supervisor would be a logical step in any ProMES system, since the primary job of a unit supervisor is to manage the human and material resources under his control in such a way that the organization's objectives are accomplished. How well the unit scores on ProMES is a good index of how well this supervisory function is being done. Consequently, the overall effectiveness score of the unit under the supervisor's control can be used as part of his performance appraisal score.

6.6.2. Design of the ProMES-portion of the management-by-results system

At the end of 1991, the following decisions were made:

* ProMES determines 30 percent of the bonus (the other 70 percent were divided among 'attainment of ISO-certificate', 'carrying out monthly activities' (after visits, service calls, etc.), and the appraisal by the field service manager.
* The region effectiveness score on the four individual indicators (MCBC, percent repeat calls, parts cost/call, labor time/call) are the basis of the system.
* Each month, only technicians of whom at least 25 hours of labor time was covered in ProMES contribute to the region effectiveness score.
* Both the absolute and the relative effectiveness scores contribute equally to the final result (comparable to the use of ProMES in the performance appraisal of the technicians).

During the measurement period (November 1991–October 1992), the ProMES system would be implemented in only six of the remaining twelve regions. With that in mind, it would not be fair to directly compare the effectiveness scores of regions working with ProMES and those for which no ProMES information was available. Therefore, it was decided to work with two groups, one consisting of the 'participation' and 'transportation' regions, the other consisting of the 'control' regions. In latter group, *no* ProMES information regarding the performance of individual technicians would be made available. In both groups, a ranking would determine the amount of bonus each supervisor would attain (setting specific absolute goals might cause the same problems that occurred in the system previously used).

According to the above specifications, ProMES was included in the supervisors' management-by-results system for the period of November 1991 through October 1992. Table 6.19 is en example of the monthly overview of region effectiveness which each supervisor in the participation/transportation group received in addition to the feedback reports of their individual technicians in his region and the region feedback report.

Table 6.19. Monthly overview of region effectiveness ('Management By Results' system of supervisors).

ProMES - Management By Results - Supervisors						region xx, month xx/92	
tech	*	MCBC	%rep	parts	labor	total	hours labor
48	n	-28	8	1	-35	-54	84
57	n	70	22	-31	-17	44	40
118	n	-17	-13	-9	-36	-75	79
137	n	7	26	23	15	71	99
199	n	52	1	8	35	96	29
229	n	-16	-52	-6	-29	-103	51
231	n	-4	70	-37	-37	-8	59
234	n	20	8	-5	-17	6	64
259	n	22	33	5	-27	33	104
296	n	20	-1	-19	-15	-15	82
298	n	-6	45	-3	28	64	42
342	n	-13	-3	-1	44	27	63
363	n	32	15	2	-18	31	87
394	n	28	28	-17	3	42	51
442	y		26	15	32	73	67
443	n	-49	27	-5	6	-21	100
Region effect.		4	16	-3	-7	10	1101
41	n		-31	-7	17	-21	12
157	n	16	70	44	74	204	5
414	n	-42	23	-3	19	-3	17
451	y						0
* = n = no time-lag, y = time lag (new technicians): MCBC result is not included							

Table 6.20 demonstrates the final result in the group consisting of the two participation regions and the six transportation regions. During the measurement period, supervisors in both groups received a monthly update of their position in the ranking.

Table 6.20. Final result group I (participation/transportation).

Group I (part/trans)		absolute		relative		overall	
rank	supervisor	score	rank	score	rank	rank	bonus
1	Enschede	29	1	+9	2	1.5	150 percent
2	Haarlem	8	5	+30	1	3.0	150 percent
3	Utrecht	20	3	+1	4	3.5	150 percent
4	Arnhem	27	2	−6	6	4.0	100 percent
5	Amersfoort	6	6	−2	5	5.5	100 percent*
6	Maastricht	−4	8	5	3	5.5	80 percent*
7	Eindhoven	9	4	−17	8	6.0	80 percent
8	Den Haag	−2	7	−7	7	7.0	80 percent

* tie-break (Amersfoort: 6−2=4, Maastricht −4+5=1)

6.6.3. Further developments

When the experiences with the management-by-results system were evaluated, supervisors and management expressed their satisfaction with the above procedure for including ProMES information. For this reason, the use of ProMES for determining the supervisors' annual bonus would be continued in the period from November 1992 to October 1993 (with some minor changes). Because implementation of the ProMES system in the remaining 'control' regions would start in the beginning of 1993, one group comprised of all 14 regions would be used instead of two groups.

The two field service managers and the product support manager agreed to the inclusion of ProMES in their pay-for-performance system (ProMES would determine 25 percent of their annual bonus). This represented a further anchoring of ProMES in the service hierarchy. In contrast to the system of the supervisors, the system of the middle managers consisted of five levels of absolute effectiveness related to four bonus-levels (below 0: 0%, 0-9: 80%, 10-17: 100%, 18-25: 125%, above 25: 150%). For the field service managers, the result would be determined by the overall effectiveness of the regions for which each of them was responsible. All regions would contribute to the product support manager's result.

In contrast to many technicians, the supervisors were reasonably satisfied with the way ProMES-information contributed to their pay-for-performance system. Several reasons can be given for this. Unlike the technicians, the supervisors had already worked with an imposed measurement system, which they did not consider accurate. Also, the ProMES system contained measures similar to those already used, the main difference being the greater validity of the ProMES measures (higher accuracy, importance weighting, type-dependent indicators). Finally, while ProMES did not account for differences in circumstances (temporary technical problems, handling of the machine by the customer, etc.), these differences were assumed to have a smaller effect on a region's effectiveness score than on a technician's effectiveness score. Finally, because the interdependence between regions was very small compared to the interdependence between technicians within a region, increased competition between supervisors would not have significant dysfunctional effects. However, one should be aware that a supervisor may be tempted to turn a blind eye to any 'gaming' of the system by some of his technicians, if he were to obtain large benefits from distortion of data.

6.7. SUMMARY

This chapter has provided a qualitative assessment of the control-loop design, based on the premiss that the effectiveness of the ProMES control loop is determined by the extent to which general design criteria and ProMES design criteria are fulfilled and by the way situational constraints are taken into account.

The results suggest that the general design criteria were partially fulfilled: the perceived amount and utility of the feedback increased dramatically in the conditions using ProMES; however, in neither condition specific goals were set. The participative design process resembled a 'joint decision-making strategy', in which the technicians' opinion had been amply taken into account. The ProMES-design essentially confirmed to the criteria, although some limitations had to be accepted regarding the completeness and controllability of the indicators and the accessibility of the feedback report. These limitations were mainly caused by two situational constraints: 'individual task/ interdependencies' and 'complexity of work flow'.

The implementation of the system by means of bilateral feedback meetings was partially successful. Whereas the majority of technicians had started discussing their feedback reports with their supervisors, the next phase (setting specific, challenging goals) had not been started. In general, supervisors carried out the introduction and the discussion of the feedback reports according to the key learning points specified in the training program. The technicians in the participation condition expressed a more positive attitude towards the accuracy and usefulness of system than their colleagues in the transportation condition. An illustrative comparison of two transportation regions suggests that the supervisor's attitude toward ProMES and his leadership behavior during the bilateral feedback meetings influences the attitude of the technicians toward the system and (to a lesser extent) the amount of productivity change.

Two by-products of the implementation of ProMES at Nashuatec were discussed: the use of ProMES-information in the performance appraisal of the technicians and in the 'Management By Results'-system of the supervisors. An interesting dilemma has been identified for the first use in the Nashuatec context: linking ProMES to the performance appraisal system is necessary to achieve consistency of control systems, but by doing so, several negative side-effects are created which impede the design and implementation of the system.

Chapter 7

General conclusions and discussion

The first objective of the research project was to determine whether the ProMES method for designing performance management systems could be successful in a context which differed in several respects from the context in which the system was tested for the first time. The second objective involved a test of the transportability of a ProMES system. The extent to which these objectives have been attained will be discussed in first section of this final chapter. Next, we will summarize the contributions of this study to the practical application of the ProMES method, as well as some issues that have not been resolved completely. This thesis will be concluded with some suggestions for future research on performance management and ProMES.

7.1. MAIN CONCLUSIONS

In this research project, a design approach was taken to the development and implementation of a ProMES system. Within this design framework, a normative 'accepted control loop model' guided the design and implementation processes. Based on the goal setting and feedback literature, the literature on ProMES, and an inventory of context characteristics, a list of requirements was drawn up. This tentative model proved to be very useful as a basis for evaluating the effectiveness of the interventions made in the design, implementation, and transportation of the system.

The generalizability of the system–a main area of interest in research on ProMES –was examined in the light of this research question: *"How can an effective ProMES control loop be designed in the Nashuatec setting with its specific characteristics?"*. In retrospect, it has become clear that the field service department of Nashuatec represented an extremely difficult context for effective performance management by means of ProMES. Several context factors were responsible for this. The complex and heterogenous work flow (caused by the different types of photocopiers) and the interdependencies that existed within an individual task environment posed several 'technical' challenges in the design phase of the project. The limited possibilities for horizontal communication necessitated a different set-up of the design procedure. The technicians' mistrust of management's intentions, the initial wait-and-see attitude adopted by management, and the possible future use of ProMES information in the performance appraisal system slowed down the design process and negatively influenced the acceptance of the system.

Notwithstanding the complex design problems that arose, participative design and implementation of the system in two field service regions caused a substantial productivity increase. Experiences during the design and implementation process in the participative condition demonstrate that designing effective solutions for these 'unique' problems is of critical importance to the effectiveness of the control loop; it is the combination of the high quality of the ProMES method and a design process in which the concrete elaboration of the performance management principles is attuned to the situation at hand that makes for a successful design.

7.1.1. Participative design and implementation

The participative design process in two service regions basically required one iteration in which the properties of the prototype system were brought into better agreement with the list of requirements of the normative control-loop model. Both the design regions and the management had a high opinion of the redesigned system: they felt that the complexities of a technician's job were very well accounted for and that the system provided the technicians with valid and useful feedback. Management, supervisors, and technicians even agreed upon a satisfactory procedure for using ProMES information in the performance appraisal.

Implementation of the redesigned system had positive effects on productivity (it caused an 18-point increase compared to the control condition). Two factors are assumed to have been responsible for this. First, the improved quality of feedback (amount and utility) increased the potential use of the feedback. Secondly, the increased perceived importance of the feedback due to its use in the individual performance appraisal of the technicians involved increased the actual use of the feedback. An additional seven-point productivity increase resulted from the institution of bilateral feedback sessions. These meetings between a supervisor and his individual technicians represented the implementation of a feedback-meeting structure suitable for the dominant individual task situation in the field service department.

Ultimately, all three (management) objectives to be attained through a ProMES system have been met in the participation condition: productivity improvement of the service technicians was achieved, the system provided valid measurement of technicians' performance and valid feedback to the technicians, and use of ProMES information contributed to a more accurate performance appraisal (distribution of the performance bonus).

7.1.2. Transportation

The second main design question of the project reads as follows: *"In this setting, can the design be successfully implemented in other–comparable–groups, without going through the participative design process?"* Whereas the participative design of ProMES in two service regions can be characterized as a 'joint decision-making strategy', the transportation of the system to six other regions reflects an 'own decision with explanation' strategy (Heller, 1971). The transportation of the ProMES system consisted of three major activities: an introductory meetings, bilateral feedback sessions (supported by a feedback and goal setting training program for the supervisors involved), and an evaluation meeting.

In Table 7.1, the participation and transportation conditions are compared with respect to productivity effects and subjective reactions to the system. The mean productivity increase in the transportation condition amounted to about half the increase obtained in the participation condition. Furthermore, subjective reactions to the system –moderately positive in the transportation condition–tended to be negative in the transportation condition. An important cause for the low degree of acceptance of the system in the transportation condition has been the proposed future use of ProMES in the technicians' performance appraisal. In the regions that designed the system, this had already caused some problems during the design and pilot-implementation of the system. Ultimately, however, a link between ProMES and the performance appraisal was accepted by the design regions, after their suggestions for improving the system had been implemented. The consistency between the performance management system and the appraisal system is assumed to be an important cause for the positive results in the participation condition. In the transportation condition, ProMES was perceived as a performance appraisal system rather than as a self-regulatory system from the beginning, which led to extremely high demands on the validity of the system (indicators should be 100 percent controllable and 'distortion-proof', etc.), illustrative of the 'incompatible tendencies'-view discussed in Section 6.5.3. Perhaps, the group effort of creating the best possible measurement and feedback system and a high level of understanding and 'ownership' are a prerequisite for having enough trust in the system and in the way colleagues use it to take up the challenge for including it in the performance appraisal.

Table 7.1. Overall comparison of the participation and the transportation condition.

	Participation	Transportation
Decision-making strategy	joint decision-making	own decision with explan.
Mean productivity change	+ 21	+ 11
Perceived validity	moderately positive	doubtful
Usefulness	moderately positive	doubtful
Link with performance appraisal	moderately positive	negative

Although transporting a ProMES system is by no means impossible, one should probably expect a lower initial degree of understanding and acceptance of the system, as well as a smaller productivity increase. The feasibility of successfully transporting a ProMES system, like designing it, is probably dependent on some situational constraints that play a part in the organization. It should be noted that a particular way of transporting a ProMES system was undertaken at Nashuatec–involving one basically unalterable system for use in all units–and that other ways of transporting a ProMES system may produce different effects.

7.1.3. Comparison across studies

In Table 7.2, the mean effects and effect-ranges from the Nashuatec study, the US Air force study (Pritchard et al., 1988, 1989) and Guzzo et al.'s (1985) meta-analysis of feedback and goal setting intervention programs are compared.

Table 7.2. Comparison of effects across studies (d statistic).

Study	Conditions (# units)	Mean *d*	Range*
ProMES: Nashuatec	Participation (2)	2.08	1.62 - 2.53
	Transportation (6)	0.82	–0.80 - 1.74
ProMES: US Air force	Feedback (5)	2.44	1.75 - 3.70
	Feedback & goal setting (5)	4.54	2.92 - 6.24
Guzzo, Jette & Katzell (1985) meta-analysis	Feedback (26)	0.35	0.08 - 0.62
	Goal setting (& feedback) (96)	0.75	0.57 - 0.93

*= Guzzo et al.'s range is a 95% confidence interval

The effects obtained in Nashuatec's participation condition are much larger than the effects found in Guzzo et al.'s (1985) meta-analysis: the smallest participation effect (d=1.62) falls outside the range found in Guzzo et al.'s goal setting studies (0.57-0.93). The mean effect in Nashuatec's transportation condition (d=0.82) is comparable to the goal setting effects found by Guzzo et al. (d=0.75). However, the variation in the Nashuatec effects is much larger; it even includes a negative effect for one of the regions. Compared to the US Air Force ProMES-interventions, the Nashuatec effects are smaller, although the mean participation effect (d=2.08) approaches the mean feedback effect found in the US Air Force program (d=2.44).

In summary, the mean effects of the ProMES interventions fall in between the mean US Air force effects and the mean Guzzo et al. effects; the participation effects are close to the US air Force feedback-effects, whereas the mean transportation effect is close to but exceeds the Guzzo et al. 'goal setting plus feedback' effect.

7.2. CONTRIBUTIONS TO THE PROMES METHOD

In the course of the design and implementation process, several modifications of and additions to the ProMES method were carried out successfully (in part necessitated by the context factors). These may prove to be an enrichment for the ProMES method and useful additions to the 'pragmatical knowledge-base' on designing and implementing ProMES systems.

The large design team necessitated the use of additional non-interactive decision-making techniques (Nominal Group Technique, Delphi questionnaires), which proved useful in generating and rating products and indicators. Scaling techniques, such as the technique of paired comparisons, were useful in determining the relative importance of a large number of indicators.

A strategic cost analysis which assessed the cost effects of standard increases of indicator values for a representative set of photocopiers, provided objective information for use in re-establishing the relative importance of the indicators. Apart from improving the link between ProMES and the company's policy regarding service, this increased confidence of the management in the system. The results from the cost analysis also provided a basis for evaluating potential financial gains of effectiveness increases as measured in the ProMES system.

The contingencies proved invaluable in accommodating for the heterogeneous and unpredictable work flow of different types of copiers. By establishing a complete set of indicators for each type of copier, a troublesome previous source of inaccuracy of measurement was solved (albeit at the expense of increased complexity of the system).

In order to retain comparability of indicator values across different types of copiers, a historical distribution of indicator values covering a one year period was used in determining the range of indicator values and the shape of the contingency function. In the feedback phase of the project, a continuously updated distribution was useful when periodically adjusting the contingencies (this was justified because the adjustments resulted from of external changes such as a decreased copy volume or additional maintenance requirements).

The ProMES method, although originally aimed at developing group performance management systems, turned out to be suited for developing an individual performance management system as well. However, the specific context–an individual task with interdependencies–caused several technical problems. Particularly, controllability and accuracy of indicators at the individual level of measurement and prevention of dysfunctional competition between technicians required special attention. Ultimately, a full-scale implementation of ProMES in the field service department involved generating no less than 270 individual feedback reports (containing results on the individual indicators), and 13 different region reports (containing results on the group indicators and aggregate results on the individual indicators) each month.

Typically, ProMES feedback meetings are periodical group meetings held shortly after the feedback report for the unit has become available. This type of meeting proved to be unsuited to the Nashuatec system, which emphasized the individual contribution to the organization through productivity measurement at the individual level (the group feedback report was largely an aggregate of the individual technicians' effectiveness scores). Therefore, a novel approach was taken: the individual ProMES reports were discussed during bilateral feedback meetings, involving a supervisor and his individual technicians. Only the group level indicators might involve group feedback meetings (even though the supervisors usually discussed individual observations with the technician in question during the bilateral meetings).

In the literature on ProMES (e.g. Pritchard et al., 1988, 1989; Pritchard, 1990), a lot of attention is being paid to the development of the measurement and feedback system. Much less is said about its implementation by means of feedback meetings, especially regarding the level of understanding of the system, leadership style, and problem-solving skills that are required of the unit supervisor. In the Nashuatec context, the supervisors largely lacked the in-depth understanding of the system and the specific skills needed to successfully conduct bilateral feedback meetings. Therefore, a goal setting and feedback training program, based on behavior modeling principles, was designed specifically for the bilateral feedback meetings. The objective of the training was to convey to the supervisor the knowledge and skills to clarify misunderstandings, deal with resistance to the program, and conducting the feedback and goal setting procedure in such a way that technicians will use ProMES as a self-regulatory system. The training and its key learning points started from the principle that a supportive 'problem-solving' style of leadership is most effective when working with ProMES.

The potential use of ProMES-information as input for a performance appraisal has proven to be a mixed blessing due to adverse reactions caused by this topic. Nevertheless, from a 'technical' point of view, ProMES could be very well incorporated into the performance appraisal of technicians and region supervisors. By using both the absolute level of effectiveness and the amount of effectiveness improvement/decrement, both the level of competency and the level of motivation of individual technicians and regions were taken into account satisfactorily.

7.3. UNRESOLVED ISSUES

Even though positive results have been accomplished at Nashuatec, some design problems have not been resolved completely yet. Some issues requiring further attention are discussed.

A general design criterion that has not been met concerns the goals in the ProMES control loop. The evaluation of the bilateral feedback meetings revealed that one of the objectives of the supervisor training program had not been achieved: using ProMES for setting specific and challenging productivity goals. To some degree, this may have been caused by the relatively short period of time the bilateral feedback procedure had been in operation and the fact that setting quantitative goals represented the final phase of this procedure and thus received limited attention. On the other hand, there are indications that some preconditions for setting specific (quantitative) and challenging goals have not been not fulfilled. Looking at the Key Learning Points from the feedback and goal setting program (Table 5.5), it is clear that generating possible causes for low and high productivity is an important step in the goal setting procedure. General cues for identifying these causes can be found in the feedback reports. Reviewing the effectiveness scores on the indicators and on the types of copiers may reveal areas in which the technician is either above or below expectations. Certain combinations of effectiveness scores (e.g. high scores on 'Labor time per call' combined with low scores on 'Mean Copies Between Calls') may direct attention to avenues for improving the overall effectiveness score. In cases such as these, the outcome feedback provided by ProMES may be sufficient to get a general idea of the causes that play a role and to implement solutions that lead to improved productivity. There is some anecdotal evidence, however, that in most cases the ProMES outcome feedback is not sufficient to identify causes and implement effective solutions and that more detailed 'process' feedback is required. This process feedback could either be direct (e.g. a technician who returns to his own repeat call and discovers that the malfunction was caused by a spare part he should have replaced; verbal feedback from a supervisor who observes the execution of a call) or postponed (e.g. a periodic overview of the malfunction causes of the machines previously repaired by the technician). These observations are in accordance with findings from the literature which demonstrate that for complex tasks, feedback and goals will affect task performance through strategy development rather than through increased effort, attention, and persistence (Wood et al., 1987). An important finding in this respect is that outcome feedback may be insufficient for developing new task strategies and that specific process feedback (information concerning the manner in which an individual implements a work strategy) may be more effective for this purpose (Earley et al., 1990). Since the direct forms of process feedback are difficult to realize–only half the calls of a technician are consecutive calls on the same machine and the supervisor visits only a few times per year–postponed process feedback may be an effective step toward improving the problem-solving process and creating a precondition for effective goal setting.

The second unresolved issue concerns the zero point as defined in the Nashuatec system. For each of the type-dependent indicators, the effectiveness score of zero corresponded to the average nation-wide indicator value on each copier included in the system. This decision was made to facilitate the development of mutually comparable contingencies for different types of copiers. As an direct consequence of this procedure, a substantial group of technicians received a monthly overall effectiveness score of less than zero (between 30 and 55 percent of the technicians, depending on the national monthly effectiveness level). During evaluation meetings and in questionnaires, it became clear that a large group of technicians considered zero effectiveness the minimum level of effectiveness to be attained. The literature on individual's responses to negative feedback predicts that the initial response by individuals who find they have not attained a standard or goal is to intensify their efforts or to change task strategies (provided the individual's self-efficacy is not low). However, repeated failures to reach the standard eventually result in abandoning the standard or turning against the feedback system (e.g. doubting its accuracy, blaming negative feedback on external factors, etc.) (Campion & Lord, 1982; Ilgen et al., 1979; Podsakoff & Farh, 1989; Taylor et al., 1984). Both responses have occurred in the Nashuatec program. During the implementation phase of the program, the main argument used by the facilitators was that the zero point was only a solution to a technical problem and that no general meaning of 'expected' or 'acceptable' should be attached to it. For a substantial group of technicians–inexperienced technicians, technicians less than average ability, or technicians working in difficult circumstances–it would not be realistic to strive for zero effectiveness and a level below zero may be expected or even above expectations. For other, highly experienced technicians, the zero point may even represent an unacceptable level of effectiveness. It was, therefore, recommended that technicians use their own past performance as a reference point, and think about ways to improve that level without taking the zero point into consideration. This line of reasoning was incorporated in the Key Learning Points for goal setting (see Table 5.5): the technician should be asked to establish his own maximum, zero point and minimum effectiveness scores. However, this is only part of the solution, since the unavailability of information about the causes of the ProMES-outcomes may decrease self-efficacy (and increase frustration) for technicians who are willing and essentially able to improve their level of effectiveness but lack a clear strategy for doing so. Thus, the importance of 'process' feedback is again demonstrated.

Finally, the use of ProMES information in the technicians' performance appraisal remains an issue of ongoing concern. Since it will not be possible to create a system which is 100 percent controllable, complete, and 'distortion-proof', using a procedure which is not optimal from a 'consistency between control systems'-viewpoint, but which is accepted by the technicians and supervisors (as discussed in Section 6.5.4) may decrease the resistance of the transportation technicians toward the system. It is conceivable that effective use of ProMES in the bilateral feedback meetings and increased utility of the system through the addition of process feedback will gradually increase the acceptance of the system in the transportation regions to a point at which a more direct and consistent link between ProMES and the individual performance appraisal would become an accepted topic of discussion.

7.4. SUGGESTIONS FOR FUTURE RESEARCH

In this final section, we will briefly discuss four topics that merit attention in future research on performance management and ProMES.

Test of the normative model. The accepted control-loop model has proven to be a useful framework for guiding and evaluating the interventions in the design and implementation of the system. Whereas some elements of the model have been very well established (e.g. the general design criteria for goal setting and feedback), other elements of the model are more tentative. It would be worthwhile to test the validity of the model (does it differentiate between successful and unsuccessful ProMES programs?) and to use this model or a similar model (e.g. Schoonen, 1993) in future ProMES projects. Some relevant questions are:

- Which context factors are dominant in which types of organizations/units/tasks?
- What is the relative importance of the criteria (in general and in a specific context)?
- How can the degree to which the criteria are met be measured? For some criteria, mostly questionnaire-type measures are suggested by Schoonen (1993). Nevertheless, it would be useful to have quantitative measures for some of the criteria. For example, the 'controllability of indicators'-criterion has remained rather elusive in the Nashuatec project. A statistical procedure for determining the controllability of indicators and identifying sources of variation (competence and effort of unit members, external factors) would be helpful.
- Which are the critical threats and success factors in designing and implementing ProMES systems?

Other modes of transportation. In this study, one particular way of transporting a ProMES system was examined: a non-participative 'own decision with explanation' strategy in which a ProMES system–designed, redesigned and implemented in two units–was implemented in six comparable units. Other ways of transporting a ProMES system to highly comparable units include:

- Developing a system with a small number of representatives from each unit, and then obtaining approval from the other units/unit members. This requires extensive horizontal communication within units, which was not feasible in the Nashuatec context.
- Using a 'consultative' strategy in the transportation units, for example by including a extra *formal* redesign step in the transportation phase. This will probably require a larger amount of time, but it may be beneficial in case there is reason to assume that the system could be improved or that other units have a different view of some of their responsibilities.

In case the units which are to use the ProMES system are not entirely comparable, other approaches to transporting the system could be taken to save time:
- In case the units have some tasks in common, but differ in others, the part of the system that covers common tasks could be retained in the transportation unit. The transportation unit would then replace the other part of the system with their own products, indicators, and contingencies (to be developed participatively).
- In case the units fulfill the same function, but do so under different circumstances (e.g. production teams who produce a similar end-product but use different machines), the transportation units could confine themselves to establishing their own appropriate set of contingencies.

Performance appraisal and rewards. Two conflicting viewpoints have been presented regarding the connection between *individual* performance management systems and performance appraisal/reward systems. On the one hand, performance will not be regulated optimally if rewards are mainly attached to performance-dimensions other than those included in the performance management system (the 'consistency between control systems view'). On the other hand, conflicting tendencies (the need for valid information versus the need to influence the exchange-relationship with the organization) are associated with both types of systems. These will hinder the development of the performance management system and require safeguards against distortion of data (the 'conflicting tendencies view').

It seems plausible that situational factors (e.g. the amount of mutual trust and respect between employees and management) determine to some extent which view has the upper hand. A similar dilemma may occur when information generated by a group-based ProMES system is to be used within an incentive system. This topic merits further research.

Task complexity. As task complexity increases, the effect of specific, difficult goals on task performance decreases (Wood et al., 1987; see Section 2.2.1). The main explanation for this finding is that new task strategies will have to be developed to insure goal achievement. The literature also suggests that feedback focusing on the behavioral processes that generate outcomes may be more beneficial than outcome feedback in case of complex tasks (e.g. Earley et al., 1990). One could thus hypothesize that ProMES systems will yield larger productivity increases for simple tasks than for complex tasks, and that systems that include behavior-oriented indicators or additional 'process' feedback are more successful for latter tasks than systems that do not possess these features. Therefore, the generalizability of the ProMES method to complex tasks would be an interesting topic of research.

Summary

The importance of an organization's human resources for achieving organizational objectives is increasingly recognized. Performance management systems take a short-term perspective to controlling employees' performance by assisting individuals, groups, and departments in regulating their daily work through performance measurement and feedback, goal setting, and incentive procedures. A method for designing performance management systems is ProMES (Productivity Measurement and Enhancement System). This method, developed by Robert D. Pritchard of Texas A&M University, is characterized by a high degree of participation of the employees in the design of the measurement system, by a decision-making process of discussion until consensus, and by a process of hierarchical review and approval. These contribute to a high degree of acceptance of the system by employees and management. The measurement system itself possesses a number of desirable characteristics, such as a single overall index of productivity and weighted subindices of productivity on a common metric. The first field test of the ProMES method with five units of an US Air Force base resulted in positive effects on productivity that were much larger than those reported in the literature on feedback and goal setting systems.

This thesis reports on the ProMES research program in the field service department of Nashuatec, a leading Dutch supplier of office equipment (photocopiers, fax machines, and laser printers). This study is one of the first in a series, conducted by the Technology and Work department of the Graduate School of Industrial Engineering and Management Science at Eindhoven University of Technology.

The study focuses on two issues. First, the generalizability of the ProMES approach that has been very successful in the US Air Force setting is tested by using it to design a performance management system in a very different setting. Secondly, the importance of a participative design procedure is assessed by implementing a system which has been designed participatively in two units in other comparable units without going through the participative design process (non-participative 'transportation' of the system). To test the effects of the interventions made, a quasi-experimental time series design with control groups is used. Within a design cycle framework, a normative model for designing a ProMES control loop is employed to guide the interventions made within the ProMES approach. In this model, three types of design criteria and constraints are distinguished. Based on the literature on feedback and goal setting, general design criteria for feedback and goals are formulated. The literature on ProMES supplies criteria which the design process, the system, and its implementation should meet. Thirdly, situational constraints, posed by the specific organizational context, have to be taken into account.

The field service department of Nashuatec, which represents the research setting, is divided into 14 regions, each consisting of approximately 20 technicians and a supervisor. The technicians' job primarily consists of visiting customers to repair machine

malfunctions (mainly photocopiers) and carry out preventive maintenance. The management of Nashuatec had three reasons for starting a ProMES program for its service technicians. First, ProMES could help maintain or improve the quality of service while reducing its cost. Second, a ProMES system would meet the technicians' recurrent request for valid and useful feedback about their performance. The third reason involved the possible future inclusion of ProMES information in the performance appraisal of the technicians.

Five situational characteristics have been identified that represent the main design problems to be solved in the design and implementation of a ProMES system at Nashuatec: 1) interdependence between technicians within an individual task; 2) high complexity of the work flow; 3) a lack of horizontal and vertical communication; 4) a top-down organizational culture and cultural differences; 5) performance appraisal and rewards issues.

The first phase of the study involved the participative design and implementation of a ProMES system in two service regions. In both regions, a design team, consisting of the technicians, their supervisor and two facilitators, worked on the system. The design process took 13 meetings in both regions (a throughput time of 16 months). Limited possibilities for horizontal communication combined with the size of the design teams necessitated the use of non-interactive decision-making techniques (Nominal Group Technique, Delphi Technique, and scaling techniques) in addition to group discussion until consensus.

A five-month pilot implementation of the system (feedback of productivity information and group feedback meetings) in the two design regions was only partially successful: although some positive results were obtained, no productivity improvements were obtained in the most important areas of responsibility.

In a participative problem analysis, the design teams identified several causes for the lack of productivity improvement. Without exception these causes referred to design criteria or context factors from the normative model that had not been fulfilled (e.g. invalidity of some indicators, a lack of visible management commitment, an unclear connection with the performance appraisal). In a discussion with management representatives, satisfactory solutions were agreed upon and incorporated into the ProMES system.

The final ProMES system consists of seven performance measures ('indicators') covering four areas of responsibility ('products'). Four indicators are type-dependent, i.e. technical characteristics of different types of copiers that influence the indicator values are taken into account through the development of separate 'contingency' functions for each type of copier included in the system, thereby providing a solution for the 'complexity of work flow'-context factor. Through these contingencies–an important element of all ProMES systems–indicator values on different indicators/types of copiers are converted into effectiveness scores that can be compared across indicators/ copiers. Furthermore, an overall index of productivity is provided. These four indicators measure the performance of individual technicians ('individual indicators'), whereas the other three indicators measure the performance of the region as a whole ('group indicators' measured by the supervisor). The interdependencies between technicians (an important context factor, since the same machine may successively be repaired by different technicians) are taken into account in two ways: the system includes indicators stimulating cooperation between technicians and it provides two levels of feedback pertaining to the performance of individual technicians and to the performance of the region as a whole. Each month, every technician receives two feedback reports: one report covering his individual productivity and one report covering the productivity of the region as a whole. In addition, graphical feedback covering a twelve-month period is provided.

A ten-month implementation of the re-designed system in the two regions caused significant and substantial productivity increases, both for the individual and the group indicators. Both management and the technicians and supervisors involved had a positive opinion of the system: it was considered much more valid and useful than any other measurement system used in the past. Also, agreement was reached on a procedure for using ProMES information in determining the annual bonus of the service technicians involved.

The second phase of the study involved the non-participative implementation (transportation) of the ProMES system in six of the remaining twelve regions. This transportation consisted of four elements. First, in each region, a standardized introductory meeting provided the technicians with a basic introduction into the ProMES system which they were to use. Secondly, follow-up to this introductory meeting was provided by means of bilateral feedback meetings between the supervisor and his individual technicians. These meetings resulted from the realization that an individualized feedback and goal setting procedure would be a prerequisite for optimum use of ProMES as a control loop in this predominantly individual task context (this was one of the recommendations from the participative problem analysis). In order to convey to the supervisors the knowledge and skills to conduct effective bilateral feedback meetings with their technicians, a feedback and goal setting training program, based on behavior modeling principles, was designed and carried out (the supervisors from the participation regions also took part in the training program and the feedback meetings). Ten months after the introductory meeting, the transportation process was concluded with an evaluation meeting.

Implementation of the system by means of bilateral feedback meetings resulted in a small additional productivity increase in the participation regions. A somewhat larger increase (significant compared to the control regions) was realized in the transportation regions. The overall effect in the participation condition was more than twice as large as the effect in the transportation condition. The overall participation effect exceeds the goal setting and feedback effects reported in the literature and approaches the ProMES feedback effect found in the US Air Force program. The transportation effect is comparable with the mean effect found in the literature.

The extent to which the design criteria and situational constraints from the normative model had been met was assessed by means of, among other things, two questionnaires. The technicians' responses to a 'Goal setting and Feedback Questionnaire' revealed that the general design criteria for feedback goals had been partially fulfilled: the perceived amount and utility of the feedback received had increased strongly (the participation technicians reported a higher utility than their colleagues in the transportation condition); however, hardly any specific goals had been set. Barriers to effective goal setting were identified, such as the low acceptance of negative effectiveness scores and the lack of information on causes of ProMES results, which hampered the development of task strategies. A questionnaire measuring the extent to which the supervisors had adhered to the Key Learning Points of the training program, which specified effective leadership behaviors during the bilateral feedback sessions, revealed limited progress of the bilateral feedback meetings. Although the first phase (presenting objectives and procedure/dealing with resistance) and the second phase (discussion of the feedback reports) had generally been carried out satisfactorily, the setting of specific, challenging goals (phase 3) had hardly been attempted.

In the participation condition, the technicians' attitudes toward the validity and usefulness of the system, the usefulness of the feedback meetings for attaining productivity improvements, and the (proposed) use of ProMES information in the performance appraisal were moderately positive. In the transportation conditions, the technicians' reactions tended to be negative, especially concerning the proposed use of ProMES information in their individual performance appraisal. Latter issue suggests that the positive effect of a high consistency between the short-term performance management system and the long-term performance appraisal system (optimum performance regulation) may be offset by negative side-effects (extremely high demands on the accuracy and controllability of the ProMES indicators, a tendency to 'game the system', and distrust towards management and colleagues).

Notwithstanding the complex design problems that arose in the Nashuatec context, the participative design and implementation of the ProMES system caused a substantial productivity increase. Experiences during the design and implementation processes in the participation condition demonstrate that designing effective solutions for 'unique' problems is of critical importance to the effectiveness of the ProMES control loop. Therefore, with regard to the generalizability of the ProMES method, the general conclusion is: an effective ProMES performance management system results from a combination of the high quality of the ProMES method and a design process in which the concrete elaboration of the performance management principles is attuned to the situational characteristics of the setting.

With regard to the transportability of ProMES systems, the general conclusion is: although transporting a ProMES system is by no means impossible, one should expect a lower initial degree of understanding and acceptance of the system, as well as a smaller productivity increase. The feasibility of successfully transporting a ProMES system, like designing it, is probably dependent on the situational constraints that play a part in the organization.

Samenvatting (summary in Dutch)

Binnen Human Resource Management is het optimaliseren van de bijdrage van medewerkers aan de organisatiedoelen via prestatiesturing een actueel onderwerp. Prestatiesturingssystemen beheersen de prestaties van medewerkers vanuit een korte termijn perspectief. Ze reguleren het dagelijkse werk van individuen, groepen en afdelingen via het meten van productiviteit, het geven van terugkoppeling, het stellen van doelen en (eventueel) het koppelen van beloningen aan gerealiseerde prestaties. Een methode voor het ontwerpen van prestatiesturingssystemen is ProMES ('Productivity Measurement and Enhancement System'). De ProMES methode, ontwikkeld door Robert D. Pritchard van de Texas A&M University, wordt gekenmerkt door een grote betrokkenheid van de medewerkers bij het ontwikkelen van het meetsysteem, een 'discussie tot consensus'-besluitvormingsmethode en een proces van afstemming in de hiërarchische lijn. Deze kenmerken dragen ertoe bij dat het systeem geaccepteerd wordt door de medewerkers en het management. Het meetsysteem zelf bezit een aantal gewenste kenmerken, zoals een totaalindex van productiviteit en gewogen subindices die in dezelfde schaal worden uitgedrukt. De eerste praktische toepassing van de ProMES methode vond plaats in vijf afdelingen van een Amerikaanse luchtmachtbasis. Dit resulteerde in productiviteitsstijgingen die veel groter waren dan die welke normaliter in de literatuur met betrekking tot terugkoppeling geven en doelen stellen worden gerapporteerd.

In dit proefschrift wordt verslag gedaan van het ProMES onderzoek in de field service afdeling van een grote Nederlandse leverancier van kantoormachines (kopieermachines, faxen en laserprinters). Dit onderzoek vond plaats in het kader van een groter onderzoeksprogramma uitgevoerd door in vakgroep Technologie en Arbeid van de faculteit Technische Bedrijfskunde van de Technische Universiteit Eindhoven.

Het promotie-onderzoek richt zich op twee onderwerpen. Ten eerste wordt de generaliseerbaarheid van de ProMES methode die zeer succesvol was bij de Amerikaanse luchtmacht onderzocht door een prestatiesturingssysteem volgens de ProMES methode te ontwerpen in een sterk afwijkende setting. Vervolgens wordt het belang van het participatieve ontwerpproces onderzocht door het systeem dat op participatieve wijze in twee groepen is ontwikkeld te implementeren in andere, sterk vergelijkbare groepen zonder het participatieve ontwerpproces opnieuw te doorlopen ('transport' van het systeem). De effecten van de interventies worden getoetst met behulp van een quasi-experimenteel tijdreeks design met controlegroepen. Met de ontwerpcyclus als uitgangspunt, wordt een normatief model voor het ontwerpen van een ProMES regelkring gebruikt als leidraad voor de interventies die binnen de ProMES benadering worden gepleegd. In dit model worden drie soorten ontwerpeisen en randvoorwaarden onderscheiden. De literatuur over terugkoppeling geven en doelen stellen levert een aantal algemene ontwerpeisen. De literatuur met betrekking tot de ProMES methode levert criteria waaraan het ontwerpproces, het ontwerp en de implementatie van het ontwerp moeten voldoen. Tenslotte dient rekening gehouden te worden met een aantal randvoorwaarden die gesteld worden door de specifieke organisatie-context.

De field service afdeling van Nashuatec (de onderzoekssetting) is verdeeld in 14 regio's die elk bestaan uit ongeveer 20 technici en een supervisor. Het dagelijkse werk van een technicus bestaat hoofdzakelijk uit het bezoeken van klanten om storingen aan (kopieer)machines te repareren en indien nodig preventief onderhoud uit te voeren. Het management van Nashuatec had drie redenen om met een ProMES programma voor haar service technici te starten. Ten eerste zou ProMES kunnen bijdragen tot een verbetering van de service kwaliteit en/of een verlaging van de kosten daarvan. Ten tweede zou men met een ProMES systeem tegemoet komen aan herhaalde vragen van de technici om valide en bruikbare prestatie-terugkoppeling. De derde reden betrof een mogelijk toekomstig gebruik van ProMES informatie in de eindejaarsbeoordeling van de service technici.

Vijf situatie-kenmerken zijn geïdentificeerd als de belangrijkste ontwerpproblemen die moeten worden opgelost bij het ontwerp en de implementatie van ProMES bij Nashuatec. Dit zijn: 1) afhankelijkheid tussen technici binnen een voornamelijk individuele taaksituatie; 2) hoge complexiteit van het werkaanbod; 3) gebrek aan horizontale en verticale communicatie; 4) top-down organisatiecultuur en cultuur-verschillen; 5) prestatiebeoordelings- en beloningsvraagstukken.

In de eerste fase van het onderzoek is op participatieve wijze een ProMES systeem ontworpen en ingevoerd in twee service regio's. In deze regio's heeft een ontwikkelteam, bestaande uit technici, supervisor en twee procesbegeleiders in 13 bijeenkomsten een ProMES systeem ontwikkeld (doorlooptijd: 16 maanden). Wegens de beperkte mogelijkheden tot horizontale communicatie en de omvang van het ontwikkelteam zijn behalve 'discussie tot consensus' niet-interactieve besluitvormingstechnieken toegepast (nl. Nominale Groepstechniek, Delphi techniek en schaaltechnieken).

Een test-implementatie van het systeem in de twee ontwerp-regio's (terugkoppeling van productiviteitsinformatie en regionale terugkoppelbijeenkomsten) was slechts gedeeltelijk succesvol. Hoewel op deelgebieden positieve resultaten werden behaald, was dit niet het geval voor de belangrijkste verantwoordelijkheidsgebieden.

In een participatieve probleemanalyse zijn door de ontwikkelteams verscheidene oorzaken vastgesteld voor het ontbreken van productiviteitsverbetering. Deze oorzaken hadden zonder uitzondering betrekking op ontwerpeisen en randvoorwaarden waaraan niet was voldaan. Voorbeelden hiervan zijn: gebrekkige validiteit van een deel van de indicatoren, gebrek aan zichtbaar management commitment aan het programma en onduidelijkheid over de relatie met beoordeling en beloning van technici. In een discussie met een vertegenwoordiging van het management is overeenstemming bereikt over de oplossingen, die vervolgens zijn opgenomen in een ProMES-herontwerp.

Het definitieve ontwerp van het ProMES systeem bestaat uit zeven prestatie-indicatoren ('indicators'), verdeeld over vier verantwoordelijkheidsgebieden ('products'), Vier indicatoren zijn model-afhankelijk, d.w.z. er wordt rekening gehouden met technische eigenschappen van de verschillende modellen copiers die beïnvloeden welke scores op de indicatoren behaald kunnen worden. Deze complexiteit van het werkaanbod is in het systeem verdisconteerd door middel van prestatiewaarderingscurven ('contingencies', een belangrijk element van elk ProMES systeem) die voor alle modellen copiers zijn vastgesteld. Door middel van de waarderingscurven kunnen indicator-scores op verschillende indicatoren/copier-modellen vertaald worden in effectiviteitswaarden die vergelijkbaar zijn voor alle indicatoren/modellen. Bovendien komt op deze manier een overall productiviteitsindex tot stand. De vier bovengenoemde indicatoren meten de productiviteit van individuele technici ('individuele indicatoren'). Voor de andere drie indicatoren was dit niet mogelijk; zij leveren scores voor de regio als geheel ('regio indicatoren', gemeten door de supervisor). Met de afhankelijkheid tussen technici (veroorzaakt doordat eenzelfde machine achtereenvolgens door verschillende technici bezocht kan worden) wordt op twee manieren rekening gehouden. Het systeem bevat indicatoren die samenwerking tussen technici stimuleren. Tevens wordt op twee niveaus terugkoppeling verstrekt: op het niveau van de individuele technicus en op het niveau van de gehele regio. Elke technicus ontvangt maandelijks een terugkoppelrapport ('feedback report') dat zijn eigen productiviteit weergeeft en een terugkoppelrapport dat de productiviteit van de regio als geheel weergeeft. In aanvulling hierop ontvangt de technicus een grafisch overzicht van zijn productiviteit in de afgelopen twaalf maanden.

Implementatie van het herontworpen systeem leidde tot significante en substantiële productiviteitsverbeteringen, zowel voor de individuele als voor de regio indicatoren. Zowel het management als de betrokken technici en supervisors hadden een positief oordeel over het systeem: ze beschouwden het als meer valide en bruikbaar dan andere meetsystemen die in het verleden waren gebruikt. Tevens werd overeenstemming bereikt over een procedure voor het gebruik van ProMES-informatie bij de bepaling van de eindejaarsbonus van de betrokken technici.

In de tweede fase van het onderzoek is het ProMES systeem op niet-participatieve wijze geïmplementeerd in zes van de twaalf overige regio's ('transport'). Dit transport van het systeem bestond uit vier onderdelen. Ten eerste is in elke regio een gestandaardiseerde introductiebijeenkomst gehouden als globale kennismaking van de technici met het ProMES-systeem waarmee zij zouden gaan werken. Bilaterale stuurgesprekken tussen de supervisors en hun individuele technici vormden een nadere uitwerking van deze introductiebijeenkomst. Deze gesprekken behoorden tot de aanbevelingen uit de participatieve probleemanalyse: een individuele procedure voor terugkoppeling geven en doelen stellen werd gezien als noodzakelijke voorwaarde voor optimaal gebruik van ProMES als regelkring in deze hoofdzakelijk individuele taaksituatie. Om de supervisors de kennis en vaardigheden bij te brengen die nodig zijn om effectieve terugkoppelgesprekken met hun technici te kunnen voeren, is een 'behavior modeling' trainingsprogramma voor het geven van terugkoppeling en het stellen van doelen ontwikkeld en uitgevoerd. Zowel de supervisors uit de 'participatie'-regio's als hun collega's uit de 'transport'-regio's hebben deelgenomen aan deze training en hebben bilaterale stuurgesprekken gevoerd. Het transport proces is afgesloten met een evaluatiebijeenkomst (tien maanden na de introductie).

Implementatie van het systeem door middel van de bilaterale stuurgesprekken leidde tot kleine additionele productiviteitsverbeteringen in de participatie-regio's. Een iets grotere productiviteitsstijging (significant ten opzichte van de controle-regio's) werd gerealiseerd in de transport-regio's. Het totale productiviteits-effect in de participatieconditie is ruim twee keer zo groot als het effect in de transport-conditie en benadert het effect van de ProMES-terugkoppeling in het onderzoek bij de Amerikaanse luchtmacht. Het geringere transport-effect is vergelijkbaar met de effecten die in de literatuur worden aangetroffen.

De mate waarin aan de ontwerpeisen en randvoorwaarden uit het normatieve model was voldaan is onderzocht door middel van (onder andere) twee enquêtes onder de service technici. De resultaten van een vragenlijst 'Doelen stellen & terugkoppeling geven' gaven aan dat voor een deel aan de algemene ontwerpeisen was voldaan: de hoeveelheid en bruikbaarheid van de terugkoppeling was door ProMES sterk verbeterd (waarbij de bruikbaarheid door de participatie-technici hoger werd gewaardeerd dan door de transport-technici). Echter, er waren nauwelijks specifieke, uitdagende doelen gesteld. Nader onderzoek wees uit dat dit onder meer veroorzaakt werd door lage acceptatie van negatieve effectiviteitsscores en een gebrek aan informatie over de oorzaken van ProMES resultaten (dit laatste bemoeilijkte de ontwikkeling van taakstrategieën). Met een tweede vragenlijst werd onderzocht in hoeverre de supervisors zich tijdens de bilaterale stuurgesprekken aan de 'Key Learning Points' (concretiseringen van effectief leiderschaps-gedrag) van het trainingsprogramma hadden gehouden. Uit de antwoorden bleek dat zowel de eerste fase van de gesprekken (uitleggen doelstelling en procedure/omgaan met weerstanden) als de tweede fase (bespreken van de terugkoppelrapporten) volgens de richtlijnen waren uitgevoerd. Echter, aan de derde fase (doelen stellen) was men nauwelijks toegekomen.

De technici in de participatie-conditie hadden een gematigd positief oordeel over de validiteit en bruikbaarheid van het systeem, het nut van de terugkoppelbijeenkomsten ter realisering van productiviteitsverbeteringen en de procedure voor het gebruik van ProMES in de eindejaarsbeoordeling. De reacties van de technici in de transport-regio's waren overwegend negatief, met name aangaande een (toekomstig) gebruik van ProMES informatie in de eindejaarsbeoordeling. Dit laatste punt geeft aan dat een positief effect van een consistente koppeling tussen het op de korte termijn gerichte prestatiesturingssysteem en het op meer lange termijn gericht beoordelingssysteem (nl. optimale regulering van prestaties) mogelijk teniet gedaan wordt door allerlei negatieve bij-effecten, zoals extreem hoge eisen aan de nauwkeurigheid en beïnvloedbaarheid van de ProMES-indicatoren, een neiging om met gegevens te frauderen, en wantrouwen jegens de bedoelingen van management en collega's.

Concluderend kan het volgende worden gesteld. Niettegenstaande de complexe ontwerpproblemen die in de Nashuatec context een rol hebben gespeeld, hebben het participatieve ontwerp en de implementatie van het ProMES systeem geleid tot een substantiële productiviteitsverbetering. Ervaringen gedurende het proces van ontwerp en implementatie in de participatie-conditie tonen aan dat het ontwerpen van effectieve oplossingen voor 'unieke' ontwerpproblemen van groot belang is voor de effectiviteit van de ProMES regelkring. Daarom kan met betrekking tot de generaliseerbaarheid van de ProMES methode geconcludeerd worden dat effectieve prestatiesturing door middel van ProMES het gevolg is van een combinatie van de hoge kwaliteit van de ProMES methode en een ontwerpproces waarin de concrete uitwerking van de prestatiesturingsprincipes afgestemd wordt op de specifieke kenmerken van de setting.

Met betrekking tot de transporteerbaarheid van ProMES systemen luidt de algemene conclusie: hoewel het transporteren van een ProMES systeem zeker niet onmogelijk is, ligt een lagere niveau van inzicht in en acceptatie van het systeem in de lijn der verwachting, evenals een minder grote productiviteitsstijging. De mate van succes bij het transporteren van een ProMES systeem lijkt, evenals bij de participatieve aanpak, beïnvloed te worden door de situationele randvoorwaarden die een rol spelen in de organisatie.

References

Algera J.A. (1990). Feedback systems in organizations. In: C.L. Cooper & I.T. Robertson (Eds.), *International review of industrial and organizational psychology* (vol. 5). Chichester: John Wiley & Sons.

Algera J.A., Janssen P.M. & Tuijl H.F.J.M. van (1992). Sturen en stimuleren van prestaties [Managing and stimulating performance]. In: F. Kluytmans & W. Vander Meeren (Eds.). *Management van human resources* [Management of human resources]. Heerlen: Open Universiteit, Deventer: Kluwer Bedrijfswetenschappen (in Dutch).

Algera J.A. & Tuijl H.F.J.M. van (1990). Feedback systems and the management of human performance in organizations. In: P.J.D. Drenth, J.A. Sergeant & R.J. Takens (Eds.), *European perspectives in psychology.* Chichester: John Wiley & Sons.

Andriessen E.J.H. & Drenth P.J.D. (1984). Leadership: theories and models. In: P.J.D. Drenth, Hk. Thierry, P.J. Willems & C.J. de Wolff (Eds.), *Handbook of work and organizational psychology.* Chichester: John Wiley & Sons.

Ashford S.J. (1986). Feedback-seeking in individual adaptation: a resource perspective. *Academy of Management Journal, 29*, 465-487.

Balcazar F., Hopkins B.L. & Suarez Y. (1986). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7* (3/4), 65-89.

Bandura A. (1977). *Social learning theory.* Englewood-Cliffs, NJ: Prentice-Hall.

Bandura A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*, 122-147.

Bandura A. & Cervone D. (1983). Self-evaluation and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology, 45*, 1017-1028.

Bass B.M. (1981). *Stogdill's handbook of leadership.* New York: Free Press.

Berkel van A. (1990). *Ontwikkeling en invoering van ProMES (Productivity Measurement and Enhancement System) bij Vandra Golfkarton BV* [Development and implementation of ProMES (Productivity Measurement and Enhancement System) at Vandra Corrugated Fibreboard]. Graduation report, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work (in Dutch).

Campbell J.P. & Campbell R.J. (1988a) (Eds.), *Productivity in organizations.* San Francisco, CA: Jossey-Bass.

Campbell J.P. & Campbell R.J. (1988b). Industrial-organizational psychology and productivity: the goodness of fit. in: J.P. Campbell & R.J. Campbell (Eds.), *Productivity in organizations.* San Francisco, CA: Jossey-Bass.

Campbell D.J. & Gingrich K.F. (1986). The interactive effects of task complexity and participation on task performance: a field experiment. *Organizational Behavior and Human Decision Processes, 38*, 162-180.

Campbell J.P. & Pritchard R.D. (1976). Motivation theory in industrial and organizational psychology. In: M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology.* Chicago: Rand-McNally.

Campion M.A. & Lord R.G. (1982). A control systems conceptualization of the goal-setting and changing process. *Organizational Behavior and Human Performance, 30,* 265-287.

Carver C.S. & Scheier M.F. (1981). *Attention and self-regulation: a control theory approach to human behavior.* New York: Springer-verlag.

Cascio W.F. (1989). *Managing human resources: productivity, quality of work life, profits* (2nd ed.). New York: McGraw-Hill.

Cohen J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Connellan T. (1978). *How to improve human performance.* Harper & Row.

Cook T.D. & Campbell D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings.* Chicago: Rand McNally.

Coolen P.R.J.J. (1990). *Een prestatiesturingssysteem voor de service technici van Nashua Nacom BV* [A performance management system for the service technicians of Nashua Nacom]. Graduation report, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work (in Dutch).

Cooper R. & Kaplan R.S. (1988). Measure costs right: make the right decisions. *Harvard Business Review*, September/October, 96-103.

Decker P.J. & Nathan B.R. (1985). *Behavior modeling training: principles and applications.* New York: Praeger.

Delbecq A.L. & Van de Ven A.H. (1971). A group process for problem identification and program planning. *Journal of Applied Behavioral Science, 7*(4), July/August.

Delbecq A.L., Van de Ven A.H. & Gustafson D.H. (1975). *Group techniques for program planning.* Glenview, IL: Scott, Foresman & Co..

Drury C. (1989). Activity Based Costing. *Management Accounting (CIMA), 67,* 60-66.

Earley P.C., Northcraft G.B., Lee C. & Lituchy T. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal, 33,* 87-105.

Edwards A.L. (1957). *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts.

Florusse L.B. & Wouters M.J.F. (1991). Ontwerpgericht onderzoek in de bedrijfskunde [Research for design in industrial engineering]. *Bedrijfskunde, 63,* 70-79 (in Dutch).

Fombrun C.J., Tichy N.M. & Devanna M.A. (1984). *Strategic human resource management.* New York: John Wiley & Sons.

Goldstein A.P. & Sorcher M. (1974). *Changing supervisor behavior.* Pergamon Press Inc.

Groot A.D. de (1970). *Methodologie: grondslagen van onderzoek en denken in de gedragswetenschappen* (5th ed.). [Methodology: foundations of research and thinking in the behavioral sciences]. 's Gravenhage: Mouton (in Dutch).

Gustafson D.H., Shukla R., Delbecq A.L. & Walster W.H. (1973). A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organizational Behavior and Human Performance, 9,* 280-291.

Guzzo R.A. (1988). Productivity research: reviewing psychological and economic perspectives. In: J.P. Campbell & R.J. Campbell (Eds.), *Productivity in organizations.* San Francisco, CA: Jossey-Bass.

Guzzo R.A., Jette R.D. & Katzell R.A. (1985). Effects of psychologically based intervention programs on worker productivity: a meta-analysis. *Personnel Psychology, 38,* 275-291.

Hackman J.R. & Oldham G.R. (1976). Motivation through the design of work: test of a theory. *Organizational Behavior and Human Performance, 16,* 122-147.

Hackman J.R. & Oldham G.R. (1980). *Work redesign.* Reading, MA: Addison-Wesley.

Hart S., Boroush M., Enk G. & Hornick W. (1985). Managing complexity through consensus mapping: technology for the structuring of group decisions. *Academy of Management Review, 10,* 587-599.

Heller F.A. (1971). *Managerial decision-making: a study of leadership styles and power-sharing among senior managers.* London: Tavistock.

Henne D. & Locke E.A. (1985). Job dissatisfaction: what are the consequences. *International Journal of Psychology, 20*, 221-240.

Ilgen D.R., Fisher C.D. & Taylor M.S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*, 349-371.

Ilgen D.R. & Klein H.J. (1988). Individual motivation and performance: cognitive influences on effort and choice. In: J.P. Campbell & R.J. Campbell (Eds.), *Productivity in organizations.* San Francisco, CA: Jossey-Bass.

In 't Veld J. (1985). *Analyse van organisatieproblemen: een toepassing van denken in systemen en processen* [Analysis of organizational problems: an application of thinking in systems and processes]. Amsterdam/Brussel: Elsevier (In Dutch).

Jacoby J., Mazursky D., Troutman T. & Kuss A. (1984). When feedback is ignored: disutility of outcome feedback. *Journal of Applied Psychology, 69*, 531-545.

Janis I. (1972). *Victims of Groupthink.* Boston: Houghton Mifflin Company.

Janssens J.M.A.M. (1988). *'-Ogen' doen onderzoek: een inleiding in de methoden van sociaal-wetenschappelijk onderzoek* ['-Ogists' do research: an introduction into the research methods of the social sciences]. Amsterdam/Lisse: Swets & Zeitlinger (in Dutch).

Janssen P.M., Tuijl H.F.J.M. van & Algera J.A. (1987). Prestatiebeoordeling in zeven organisaties [Performance appraisal in seven organizations]. *Gids voor personeelsmanagement, 5*, 16-18 (in Dutch).

Judson A.S. (1982). The awkward truth about productivity. *Harvard Business Review,* September/October, 93-97.

Kendrick J.W. (1984). *Improving company productivity.* Baltimore: Johns Hopkins University Press.

Klein H.J. (1989). An integrated control theory model of work motivation. *Academy of Management Review, 14*, 150-172.

Kleingeld P.A.M. (1990). *Ontwikkeling van een prestatiesturingssysteem voor de service technici van Nashua Nederland BV* [Development of a performance management system for the service technicians of Nashua Netherlands]. Graduation Report, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work (in Dutch).

Kleingeld P.A.M. & Tuijl H.F.J.M. van (1992). *ProMES documentatiemap Nashuatec* [ProMES-manual Nashuatec]. Eindhoven University of Technology/Nashuatec. February 1992 (updated in December 1992 and February 1993) (in Dutch).

Kleingeld P.A.M. & Tuijl H.F.J.M. van (1992). *Individual and group productivity enhancement in a field service department.* Internal report, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work, July 1992.

Kleingeld P.A.M. & Tuijl H.F.J.M. van (in press). Individual and group productivity enhancement in a field service setting. In: Pritchard R.D. (Ed.), *Productivity improvement strategies and applications: case studies in organizations.* New York: Praeger.

Kopelman R.E. (1982). Improving productivity through objective feedback: a review of the evidence. *National Productivity Review, 2*, 43-55.

Kopelman R.E. (1986). Objective feedback. In E.A. Locke (Ed.), *Generalizing from laboratory to field settings.* Lexington, MA: Lexington Books.

Lamberts B. (1991). *Strategische kostenanalyse van het ProMES prestatiebesturingssysteem voor de technici van de afdeling service bij Gestetner Nashua* [Strategic cost analysis of the ProMES performance management system for the service technicians of the service department at Gestetner Nashua]. Graduation Report, Erasmus University Rotterdam, Department of Business Economics (in Dutch).

Latané B., Williams K. & Harkins S. (1979). Many hands make light the work: the causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

Latham G.P., Erez M. & Locke E.A. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: application to the Erez-Latham dispute regarding participation in goal setting. *Journal of Applied Psychology (Monograph), 73*, 753-772.

Latham G.P. & Locke E.A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 46*, 212-247.

Latham G.P., Mitchell T.R. & Dossett D.L. (1978). Importance of participative goal setting and anticipated rewards on goal difficulty and job performance. *Journal of Applied Psychology, 63*, 163-171.

Latham G.P. & Saari L.M. (1979a). Importance of supportive relationships in goal setting. *Journal of Applied Psychology, 64*, 151-156.

Latham G.P. & Saari L.M. (1979b). Application of social-learning theory to training supervisors through behavior modeling. *Journal of Applied Psychology, 64*, 239-246.

Latham G.P. & Steele T.P. (1983). The motivational effects of participation versus goal setting on performance. *Academy of Management Journal, 26*, 406-417.

Latham G.P. & Wexley K.N. (1981). *Increasing productivity through performance appraisal.* Reading, MA: Addison-Wesley.

Latham G.P. & Yukl G.A. (1975), Assigned versus participative goal setting with educated and uneducated woods workers. *Journal of Applied Psychology, 60*, 299-302.

Law D.E. (1975). Managing for productivity. *The Arthur Young Journal,* Summer/Autumn, 2-13.

Linstone H.A. & Turoff M. (1975). *The Delphi method: techniques and applications.* Reading, MA: Addison-Wesley.

Locke E.A., Chah D.O., Harrison S. & Lustgarten N. (1989). Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Decision Processes, 43*, 270-287.

Locke E.A. & Latham G.P. (1990). *A theory of goal setting & task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Locke E.A. & Schweiger D.M. (1979). Participation in decision making: one more look. In: B. Staw (Ed.), *Research in Organizational Behavior* (vol. 1). Greenwich, CT: JAI Press.

Locke E.A., Schweiger D.M. & Latham G.P. (1986). Participation in decision-making: when should it be used? *Organizational Dynamics, 14* (3), 65-79.

Luthans F. & Kreitner R. (1975). *Organizational behavior modification.* Glenview, IL: Scott, Foresman & Co.

MacLean B.P. (1990). Compensation: value-added pay beats traditional merit programs. *Personnel Journal*, September, 46-52.

Mahler J.G. (1987). Structured decision making in public organizations. *Public Administration Review*, July/August, 336-342.

Mahoney T.A. (1988). Productivity defined: the relativity of efficiency, effectiveness, and change. In: J.C. Campbell & T.R. Campbell (Eds.), *Productivity in organizations.* San Francisco, CA: Jossey-Bass.

Mali P. (1978). *Improving total productivity.* New York: John Wiley & Sons.

Matsui T., Kakuyama T. & Onglatco L.U. (1987). Effects of goals and feedback on performance in groups. *Journal of Applied Psychology, 72*, 407-415.

Matsui T., Okada A. & Inoshita O. (1983). Mechanisms of feedback affecting task performance. *Organizational Behavior and Human Performance, 31*, 114-122.

Matsui T., Okada A. & Kakuyama T. (1982). Influence of achievement need on goal setting, performance, and feedback effectiveness. *Journal of Applied Psychology, 67*, 645-648.

Miller K.I. & Monge P.R. (1986). Participation, satisfaction, and productivity: a meta-analytic review. *Academy of Management Journal, 29*, 727-753.

Mintzberg H. (1979). *The structure of organizations.* Englewood Cliffs, NJ: Prentice-Hall.

Mitchell T.R. (1973). Motivation and participation: an integration. *Academy of Management Journal, 16*, 670-679.

Mitchell T.R. & Silver W.S. (1990). Individual and group goals when workers are interdependent: effects on task strategies and performance. *Journal of Applied Psychology, 75*, 185-193.

Nadler D.A. (1979). The effects of feedback on task group behavior: a review of the literature. *Organizational Behavior and Human Performance, 23*, 309-328.

Nathan B.R., Mohrman A.M. & Milliman J. (1991). Interpersonal relationships as a context for the effects of appraisal interviews on performance and satisfaction: a longitudinal study. *Academy of Management Journal, 34*, 352-369.

Paquin A.R., Jones S.D. & Roth P.L. (1992). *The Hawthorne effect when evaluating productivity gains.* Paper presented at the 7th annual conference of the Society for Industrial and Organizational Psychology, Montreal, Canada, April 1992.

Podsakoff P.M. & Farh J. (1989). Effects of feedback sign and credibility on goal setting and task performance. *Organizational Behavior and Human Decision Processes, 44*, 45-67.

Pritchard R.D. (1990). *Measuring and improving organizational productivity: a practical guide.* New York: Praeger.

Pritchard R.D. (1992). Organizational productivity. In: M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., vol. 3). Palo Alto, CA: Consulting Psychologists Press.

Pritchard R.D., Algera J.A., Janssen P.M. & van Tuijl H.F.J.M. (1990). Productivity measurement and enhancement: a European perspective. Eindhoven University of Technology, unpublished manuscript.

Pritchard R.D., Jones S.D., Roth P.L., Stuebing K.K. & Ekeberg, S.E. (1986). Organizational productivity measurement: the development and evaluation of an integrated approach. *Air Force Human Resources Laboratory Technical Report,* AFHRL-TR-86-64.

Pritchard R.D., Jones S.D., Roth P.L., Stuebing K.K. & Ekeberg, S.E. (1987). The feedback, goal-setting, and incentives effects on organizational productivity. *Air Force Human Resources Laboratory Technical Report,* AFHRL-TR-87-3.

Pritchard R.D., Jones S.E., Roth P.L., Stuebing K.K. & Ekeberg S.E. (1988). The effects of feedback, goal setting and incentives on organizational productivity. *Journal of Applied Psychology Monograph, 73*, 337-358.

Pritchard R.D., Jones S.E., Roth P.L., Stuebing K.K. & Ekeberg S.E. (1989). The evaluation of an integrated approach to measuring organizational productivity. *Personnel Psychology, 42*, 69-115.

Pritchard R.D. & Roth P.J. (1991). Accounting for non-linear utility functions in composite measures of productivity and performance. *Organizational Behavior and Human Decision Processes, 50*, 341-359.

Putten J.A.M. van der (1988). *Goal setting-feedback als onderdeel van het door Nashua Nederland BV te voeren personeelsbeleid voor zijn service-technici* [Goal setting-feedback as a part of Nashua Netherlands' personnel policy for its service technicians]. Graduation Report, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work (in Dutch).

Raaijmakers J.G.W. (1993). *De psycholoog als ingenieur: over toegepast onderzoek in de psychologische functieleer* [The psychologist as engineer: on applied research in experimental psychology]. Inaugural lecture, September 23, 1993, University of Amsterdam (in Dutch).

Ridderbos A. (1992). *Selection by simulation: a work sample approach to the selection of process operators.* PhD. thesis. Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work.

Riggs J.L. & Felix G.H. (1983). *Productivity by objectives.* Englewood Cliffs, NJ: Prentice-Hall.

Rodgers R. & Hunter J.E. (1991). Impact of management by objectives on organizational productivity. *Journal of Applied Psychology, 76*, 322-336.

Roe R.A. (1989). Designing selection procedures. In: P. Herriot (Ed.), *Assessment and selection in organizations: methods and practice for recruitment and appraisal.* Chichester: John Wiley & Sons.

Rohrbaugh J. (1979). Improving the quality of group judgement: social judgement analysis and the Delphi technique. *Organizational Behavior and Human Performance, 24*, 73-92.

Rohrbaugh J. (1981). Improving the quality of group judgement: social judgement analysis and the nominal group technique. *Organizational Behavior and Human Performance, 28*, 272-288.

Roozenburg N.F.M. & Eekels J. (1991). *Produktontwerpen, structuur en methoden* [Product design, structure and methods]. Utrecht: Lemma (in Dutch).

Rosenbaum M.E., Moore D.L., Cotton J.L., Cook M.S., Hieser R.A., Shovar M.N., Gray M.J. (1980). Group productivity and process: pure and mixed reward structures and task interdependence. *Journal of Personality and Social Psychology, 39*, 626-642.

Runnion A., Watson J.D. & McWorther J. (1978). The effects of feedback and reinforcement on truck turnaround time and materials transportation. *Journal of organizational Behavior Management, 1*, 110-117.

Scheflen K.C., Lawler E.E. & Hackman J.R. (1971). Long-term impact of employee participation in the development of pay incentive plans: a field experiment revisited. *Journal of Applied Psychology, 55*, 182-186.

Schoonen M. (1993). *Design and evaluation of performance measurement systems.* Graduation report, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work.

Schweiger D.M. & Leana C.R. (1986). Participation in decision making. In: E.A. Locke (Ed.), *Generalizing from Laboratory to Field Settings.* Lexington, MA: Lexington Books.

Schweiger D., Sandberg W. & Ragan J. (1986). Group approaches for improving strategic decision making: a comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *Academy of Management Journal, 29*, 51-71.

Siegel L. & Lane I.M. (1982). *Personnel and organizational psychology.* Homewood, IL: Richard D. Irwin Inc.

Sink D.S. (1983). Multi-criteria performance/productivity measurement technique. *Productivity Management, 4*, 1-4.

Sink S.D. (1985). *Productivity management: planning, measurement, and evaluation, control and improvement.* New York: John Wiley & Sons.

Steele Johnson D., Perlow R. & Pieper K.F. (1993). Differences in task performance as a function of type of feedback: learning-oriented versus performance-oriented feedback. *Journal of Applied Social Psychology, 23*, 303-320.

Taylor M.S., Fisher C.D. & Ilgen D.R. (1984). Individuals' reactions to performance feedback in organizations: a control theory perspective. In: K.M. Rowland & G.R. Ferris (Eds.), *Research in personnel and human resources management* (vol 2). Greenwich, CN: JAI Press.

Thierry Hk. (1987). De effektiviteit van prestatiebeloning [The effectiveness of pay-for-performance]. *Economisch Statistische Berichten,* 2-9-87, 804-809 (in Dutch).

Tichy N.M., Fombrun C.J. & Devanna M.A. (1982). Strategic human resource management. *Sloan Management Review,* winter, 47-61.

Tuijl H.F.J.M. van (1990). *Toepassing van ProMES in de service afdeling van Nashua Nederland BV* [Application of ProMES to the service department of Nashua Netherlands]. Paper presented at the Euroforum conference on performance management and productivity, Rotterdam, May 31, 1990 (in Dutch).

Tuijl H.F.J.M. van, Janssen P.M., & Algera J.A. (1987). *Prestatiebeoordeling in zeven organisaties* [Performance appraisal in seven organizations]. Report EUT/BDK/24, Eindhoven University of Technology, Graduate School of Industrial Engineering and Management Science, Department of Technology and Work (in Dutch).

Tuijl H.F.J.M. van & Kleingeld P.A.M. (1992). *ProMES supervisortraining Nashuatec: feedback meeting procedure, key learning points en rol-instructies* [ProMES supervisor training program: feedback meeting procedure, key learning points and role instructions]. Eindhoven University of Technology, March 1992 (in Dutch.).

Tuttle T.C. (1981). *Productivity measurement methods: classification, critique, and implications for the US Air Force* (AFHRL-TR-81-9). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Tuttle T.C. (1983). Organizational productivity: a challenge for psychologists. *American Psychologist, 38,* 479-486.

Tuttle T.C. & Weaver C.N. (1986a). *Methodology for generating efficiency and effectiveness measures (MGEEM): a guide for commanders, managers, and supervisors* (AFHRL TP-86-26), Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Tuttle T.C. & Weaver C.N. (1986b). *Methodology for generating efficiency and effectiveness measures (MGEEM): a guide for Air Force measurement facilitators* (AFHRL TP-86-36), Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Van de Ven A.H. & Delbecq A.L. (1971). Nominal versus interacting group processes for committee decision-making effectiveness. *Academy of Management Journal, 14,* 203-212.

Van de Ven A.H. (1974). *Group decision-making and effectiveness, an experimental study.* Kent, OH: Kent State University.

Video Arts (1988). *From no to yes–the constructive route to agreement.* Videotape (25 mins).

Vlist R. van der (1990), Participatief management, ondanks kritiek van Locke en Schweiger, een aan te bevelen strategie [Participative management, notwithstanding criticism of Locke and Schweiger, a recommended strategy], *Gedrag en Organisatie, 3,* 332-343 (in Dutch).

Vosselman E.G.J. (1993). *De betekenis van de bedrijfskundige gevalstudie voor theorievorming* [The significance of the case study in management science for theory development]. Paper presented at the sixth NOBO research conference, Rotterdam, November 10, 1993 (in Dutch).

Vroom V.H. (1960). *Some personality determinants of the effects of participation.* Englewood Cliffs. N.J.: Prentice-Hall.

Vroom V.H. & Yago A.C. (1978). On the validity of the Vroom-Yetton model. *Journal of Applied Psychology, 63*, 151-162.

Vroom V.H. & Yetton P.W. (1973), *Leadership and decision-making.* Pittsburgh, PA: University of Pittsburgh Press.

Wagner J.A. & Gooding R.Z. (1987). Shared influence and organizational behavior: a meta-analysis of situational variables expected to moderate participation-outcome relationships, *Academy of Management Journal, 30*, 524-541.

Wilson R.W., Neely A.D. & Aggarwal N. (1993). Allowing for the human element: human factors in small manufacturing enterprises. *International Journal of Human Factors in Manufacturing, 3*, 193-205.

Wood R.E. (1986). Task complexity: definition of the construct. *Organizational Behavior and Human Decision Processes, 37*, 60-82.

Wood R.E. & Locke E.A. (1990). Goal setting and strategy effects on complex tasks. In: B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (vol. 12). Greenwich, CT: JAI Press.

Wood R.E., Mento A.J. & Locke E.A. (1987). Task complexity as a moderator of goal effects: a meta-analysis. *Journal of Applied Psychology, 72*, 416-425.

Appendix A
Checklists used for the group indicators

* Accuracy of History Card	: after visit
* Compliance with preventive maintenance procedures	: after visit & during call
* Correctness of behavior	: during call

CHECKLIST HISTORY CARD

Date of inspection:			Type:	
Date call technician:			Serial number:	
Technician id. number:				

		score obtained	minimum score		
1.	Copy counter		-4		
2.	Date		-4		
3.	Technician id. number		-2		
4.	Malfunction code		-4		
5.	Customer information		-1		
6.	Maintenance type		-8		
7.	Information (reverse side)		-8		-31
8.	Parts replacement (lifetime/failure)		-10	n/a*	
9.	Additional procedure		-3	n/a	
10.	Modification label		-3	n/a	
11.	Service counters		-8	n/a	
	Overall obtained score/minimum score				

* n/a = not applicable, i.e. the entire element is excluded from the determination of the overall score

Comments:

CHECKLIST PREVENTIVE MAINTENANCE AFTER VISIT

Date after visit:		Type:	
Date call technician:		Serial number:	
Technician id. number:		Maintenance type:	

(tick the appropriate box)

a. Exterior	satisfactory		1
	unsatisfactory		0
	n/a		
b. Optics	satisfactory		1
	unsatisfactory		0
	n/a		
c. Paper feed	satisfactory		1
	unsatisfactory		0
	n/a		
d. Operation test	satisfactory		1
	unsatisfactory		0
	n/a		
e. Coronas	satisfactory		1
	unsatisfactory		0
	n/a		
f. Transport	satisfactory		1
	unsatisfactory		0
	n/a		
g. Fuser and exit	satisfactory		1
	unsatisfactory		0
	n/a		
h. General impression interior	satisfactory		1
	unsatisfactory		0
	n/a		
i. Optionals (df/sorter/duplex/...)	satisfactory		1
	unsatisfactory		0
	n/a		

Number of elements "n/a":	
Number of elements applicable (9- # n/a):	
Score (number of elements "1'):	

Comments:

CHECKLIST PREVENTIVE MAINTENANCE DURING CALL

Date:		Type:	
Technician id. number:		Serial number:	
		Maintenance type:	

(tick the appropriate box)

a. Percentage of maintenance elements carried out by the technician: (to be determined with accompanying checklist)

100 %		8
81-99 %		4
<= 80 %		0

b. The technician routinely followed the work routing prescribed by the maintenance procedure:

entirely according to procedure		5
slight deviation from the procedure		$2\frac{1}{2}$
extensive deviation from the procedure		0

c. In view of the local circumstances, the technician follows the maintenance constraints prescribed by the maintenance procedure:

entirely according to procedure		3
slight deviation from the procedure		$1\frac{1}{2}$
extensive deviation form the procedure		0

d. The technician uses the right tools and cleaning materials in the prescribed way:

entirely according to procedure		4
slight deviation from the procedure		2
extensive deviation from the procedure		0

Score (sum of scores a,b,c,d): ______

Checklist 'percentage of maintenance element carried out'
(during call, section a)

Choose the appropriate alternative: 'yes', carried out; 'no', not carried out. The alternative 'n/a' (not applicable) should be chosen if the element is not part of the maintenance type in question (OS 0,1,2).

ELEMENTS	yes	no	n/a
Drum section			
Coronas			
Cleaning unit			
Development unit			
Transport			
Optics			
Paper feed			
Exterior			
Operation test			
Fuser and exit			
Rear (drive)			
Optionals (sorter/df/duplex)			

$$\% \text{ carried out} = \frac{\# \text{ 'yes'}}{12 - \# \text{ 'n/a'}} * 100\% =$$ ______ -> section a

Comments:

CHECKLIST AMBASSADORSHIP (DURING CALL)

Date: [ ] Technician id. number: [ ]

(tick the appropriate box)

a. The technician park his car in such a way as to cause no annoyance.

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

b. The technician announced his arrival and his departure with to the receptionist/door-keeper.

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

c. On arrival, the technician reported to the key operator and discussed the malfunction with him/her.

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

d. The technician places toolbox, tools, and spare parts in such a way that third parties are not hindered in their duties (taking into account the situation at the location).

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

e. The technician has taken the protective measures prescribed by the maintenance procedure (mats, plastic bags, etc.)

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

f. After completing the repair, the technician reported to the key operator, informed him/her about the malfunction and repair/ maintenance activities performed, and (if applicable) which further activities will be undertaken.

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

h. The technician cleared away any litter caused by his work.

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

i. General assessment of the get-up of the technician's car

Excellent: better than could be expected	[ ]	1
Good: car is reasonably clean	[ ]	1
not acceptable: car has not been cleaned for a long time	[ ]	0

j. General assessment of condition of the technician's tool-box, parts-box, and manuals

Excellent: better than could be expected	[ ]	1
Good: in reasonable condition	[ ]	1
not acceptable: disorderly and not well-kept	[ ]	0

k. The technician was dressed in the prescribed company clothing.

yes	[ ]	1
no	[ ]	0
n/a	[ ]	

l. General assessment of the technician's behavior/appearance/courtesy.

excellent	[ ]	1
good	[ ]	0
unaccept.	[ ]	

n/a [ ] # elements (12 - # n/a) [ ] score (# "1") [ ]

Comments:

Appendix B

Feedback reports (pilot feedback period)

The individual feedback report. In the pilot feedback period, the individual pilot feedback report consisted of three pages:

1) a summary overview of individual effectiveness (Figure 1);
2) a detailed overview per type of copier of individual effectiveness (type-dependent indicators, Figure 2);
3) a detailed overview of individual effectiveness on the type-independent indicators (two individual and three group indicators, Figure 3).

PROMES INDIVIDUAL FEEDBACK REPORT

Region : 1
Technician : Isaak N. Ventive
Period : January 1991

SUMMARY OVERVIEW OF INDIVIDUAL EFFECTIVENESS

	month	moving average	
QUALITY			
Mean Copies Between Calls	28	10	(80-,100)
Percentage repeat calls	26	2	(75-, 80)
Preventive Maintenance	18-	2-	(65-, 40)
	-----	-----	
	36	10	
COST			
Perc. return parts car stock	16	15	(80-, 80)
Parts cost per call	10-	5	(55-, 55)
Labor time per call	6-	4-	(60-, 60)
	-----	-----	
	0	16	
ADMINISTRATION			
Accuracy of History Card	5	7-	(25-, 25)
Completeness of claims		12-	(30-, 30)
	-----	-----	
	5	19-	
AMBASSADORSHIP			
Correctness of behavior	26	13	(30-, 25)
ATTENDANCE			
Percentage of capacity used	32-	10	(80-, 30)
Overall individual effectiveness	35	18	
Perc. of maximum effectiveness	7.1%	3.4%	
PRODUCT KNOWLEDGE			
Perc. return calls knowledge	0.3%	0.4%	

Figure 1. Individual pilot feedback report page 1: Summary overview of individual effectiveness.

DETAILED OVERVIEW OF INDIVIDUAL EFFECTIVENESS
TYPE-DEPENDENT INDICATORS

Region : 1
Technician : Isaak N. Ventive
Period : January 1991

	Month calls	Month value	Month effect.	Moving Average value	Moving Average effect.	calls
QUALITY						
Mean Copies Between Calls						
Type P [4100]	28	13.1 k	25	11.9 k	20	135
Type Q	1			19.4 k	12	13
Type X	14	14.9 k	1-	9.4 k	32-	85
Type Z	11	50.5 k	73	42.7 k	34	89
			28		10	
Percentage repeat calls						
Type P [4100]	33	6.1 %	15	8.4 %	3-	155
Type Q	4			6.7 %	25	15
Type X	7	28.6 %	59-	18.5 %	19-	81
Type Z	17	0.0 %	80	7.9 %	25	90
			26		2	
COST						
Perc. return calls parts car stock						
Type P [4100]		3.0 %	40	3.9 %	38	
Type Q				0.0 %	80	
Type X		14.3 %	64-	4.9 %	5	
Type Z		5.9 %	1	8.9 %	25-	
			16		15	
Parts cost per call						
Type P [4100]		28.50.-	1	21.60.-	14	
Type Q				30.10.-	7-	
Type X		9.30.-	26	17.50,-	6	
Type Z		147.50.-	46-	67.00.-	8-	
			10-		5	
Labor time per call						
Type P [4100]		93 min	10-	86 min	5-	
Type Q				105 min	32-	
Type X		81 min	22	95 min	4	
Type Z		125 min	10-	120 min	6-	
			6-		4-	

Figure 2. Pilot individual feedback report page 2: Detailed overview of individual effectiveness (type-dependent indicators).

DETAILED OVERVIEW OF INDIVIDUAL EFFECTIVENESS
TYPE-INDEPENDENT INDICATORS

Region : 1
Technician : Isaak N. Ventive
Period : January 1991

MEASURED AT INDIVIDUAL LEVEL	Month value	Month effect.	Moving Average value	Moving Average effect.
ADMINISTRATION				
Completeness of claims			50.3 %	12-
ATTENDANCE				
Percentage of capacity used	99.0 %	32-	100.8 %	10

MEASURED AT REGION LEVEL	Month value	Month effect	Moving Average value	Moving Average effect.
QUALITY				
Preventive maintenance	87.3 %	18-	89.8 %	2-
ADMINISTRATION				
Accuracy of History Card	96.0 %	5	90.8 %	7-
AMBASSADORSHIP				
Correctness of behavior	98.7 %	26	95.0 %	13

Figure 3. Individual pilot feedback report page 3: Detailed overview of individual effectiveness (type-independent indicators).

The region feedback report. In the pilot feedback period, the region feedback report consisted of two pages:

1) a summary overview of region effectiveness;
2) a detailed overview of region effectiveness on the type-independent indicators (two individual and three group indicators).

The lay-out of these reports was identical to the lay-out of the corresponding individual reports (Figures 1 and 3, respectively).

Appendix C

Establishing the relative importance of the indicators (cost-component)

The strategic cost analysis according to the Activity Based Costing method consisted of three steps (Lamberts, 1991):

1) Determining the direct and indirect activities within the service department which are influenced by ProMES.
2) Allocating the costs to activities, based on the cost drivers identified (i.e. the causal relations between the resources and the activities that consume these resources).
3) Allocating the costs to the products selected, based on use of the activities by the products.

Based on interviews with department heads within the service organization, the activities that are carried out were analyzed and their relation to ProMES was determined. The activities that would change as a result of changes in the ProMES-scores obtained were selected. Cost drivers were identified and used to allocate the costs (collected for a representative one-month period) to the activities. The costs were then allocated to three types of copiers, each representing a product segment (low, middle, and high volume). Table 1 shows the relevant departments and activities, the cost drivers identified, and the relation to some of the indicators.

Table 1. Departments, activities, cost drivers, related to indicators.

Departments	Cost drivers	mcbc	%rep	parts	labor	%ret
Service reception * recording machine and customer information	# malfunction phonecalls	X	X			
Planning * allocating technician-visits to customers	# visits	X	X			X
Field Service * solving machine malfunctions	# visits	X	X	X	X	X
Product Support * training programs * techn. support (phone) * various (all 40%)	# visits (#technicians/type) (# old vs new types)	X	X			X
Logistics * ordering * picking and packing * distribution	 # spare parts used # sp. parts ordered # return calls			 X X		 X X

Determining the cost-effects of ProMES-indicators involved the following additional steps (for each of the products):

1) Establishing comparable increases on each of the indicators; the increases were made comparable across indicators by using a standard increase of 10 percent of the range of indicator scores as defined in the contingencies[1].
2) Determining the change in activities resulting from the increased indicator scores.
3) Determining the corresponding change in costs.
4) Calculating the overall change in costs (across products).
5) Comparing the cost changes for each of the indicators.

The results of this procedure are shown in Table 2. In order to get a complete estimate of the relative importance of the indicators, the relative importance from a cost point-of-view was combined with subjective estimates of customer satisfaction (see Section 4.7.6). The cost data were also used as the basis for assessing the financial effects of the implementation of ProMES (see Section 5.4).

Table 2. Absolute and relative cost reductions.

Indicator/type	Type 4100 (low)		Type 5130 (middle)		Type 7150 (high)		Overall relative importance
result	cost	import.	cost	import.	cost	import.	
MCBC	11215	*100 %*	9055	97 %	7446	84 %	100 %
% repeat calls	2156	19 %	3233	35 %	2678	28 %	30 %
parts cost/call	6930	62 %	9370	*100 %*	6022	68 %	82 %
labor time/call	8181	86 %	8071	86 %	8907	*100 %*	96 %
% return calls	1313	12 %	1722	18 %	1375	15 %	16 %
% capacity used	2636	23 %	2192	24 %	1456	19 %	22 %
% claims	137	1 %	114	1 %	76	1 %	1 %

[1] The range from lower inflection point to upper inflection point was chosen (if available), since that would constitute a more realistic range than the range from minimum to maximum.

Appendix D

Goal setting and Feedback Questionnaire & Feedback Meeting Questionnaire: scales used

For the Goal setting and Feedback Questionnaire (GFQ), which contained 77 five-point-Likert items based in part on Algera & Van Tuijl (1990), a complete set of scales (including Cronbach's Alpha) and individual items is given. The Feedback Meeting Questionnaire, developed specifically for this project, contained 53 questions in different formats. Therefore, mostly separate items were used; two scales were constructed.

Goal setting and Feedback Questionnaire: scales	number of items	alpha
Goal specificity (results)	4	.755
Goal specificity (work strategies)	1	
Participation in goal setting	3	.766
Goal difficulty	4	.800
Goal attainability	4	.793
Knowledge of strategies	4	.795
Knowledge of priorities	4	.833
Controllability of results	6	.717
Possession of knowledge and skills needed	1	
Availability of means needed	1	
Work-related communication with supervisor	2	.731
Dependence on colleagues	1	
Dependence on supervisor	1	
Influence on colleagues	1	
Influence on supervisor	1	
Measurement of results	5	.843
Familiarity with measurement of results	1	
Amount of feedback about results	4	.914
Amount of feedback about work strategies	1	
Feedback from supervisor	1	
Feedback from colleagues	1	

GFQ: scales (continued)	number of items	alpha
Feedback from other departments	1	
Feedback from regular reports	1	
Feedback from the task itself	1	
Feedback utility	6	.880
Negative feedback	1	
Positive feedback	1	
Oral feedback	1	
Written feedback	1	
Positive reinforcement	3	.670
Negative reinforcement/punishment	3	.713
Relationship with supervisor	7	.937

Feedback Meeting Questionnaire: scales	number of items	alpha
Quality of introduction (phase 1)	5	.776
Quality of feedback (phase 2)	6	.782

About the Author

P.A.M (Ad) Kleingeld was born on June 15, 1967 in Utrecht, the Netherlands. In 1985, he received his Gymnasium-β diploma from Hertog Jan College in Valkenswaard, after which he started his study of industrial engineering and management science at Eindhoven University of Technology. He graduated cum laude from the master's program with a major in personnel management. From March 1990 to April 1994, he took part in the PhD program at the Graduate School of Industrial Engineering and Management Science as a member of the Technology and Work department. This PhD-thesis reports on the research conducted during this period. The research project focused on the design and transportation of a performance management system in the field service department of a large supplier of photocopiers. In the course of the research project he presented papers at several national and international symposia and (co-)wrote publications pertaining to the ProMES method for designing performance management systems and to its application in the framework of the PhD-project. Ad Kleingeld is currently employed at the department of Technology and Work.

May, 1994

Stellingen

behorende bij het proefschrift

PERFORMANCE MANAGEMENT IN A FIELD SERVICE DEPARTMENT

Design and Transportation of a Productivity Measurement and Enhancement System (ProMES)

van

P.A.M. Kleingeld

13 september 1994

I De gebruiker van ProMES dient zich te realiseren dat juist de toepassing van algemene principes in een specifieke situatie het creatieve ontwerpende element bevat dat in belangrijke mate bepaalt of effectieve prestatiesturing gerealiseerd wordt.

- Dit proefschrift.

II Bij het ontwerpen van een ProMES prestatiesturingssysteem leidt een taaksituatie die gekenmerkt wordt door een beperkte mate van interdependentie tussen groepsleden tot grotere ontwerpproblemen dan situaties die gekenmerkt worden door sterke interdependentie tussen groepsleden of door volledige independentie van groepsleden.

- Dit proefschrift.
- Pritchard R.D. (Ed.) (in druk). *Productivity improvement strategies and applications: case studies in organizations.* New York: Praeger.

III Bij complexe taken verdient het aanbeveling 'proces' feedback te verstrekken in aanvulling op (ProMES-) resultaat feedback; dit is met name het geval indien feedback niet door het werk zelf wordt gegenereerd.

- Earley P.C., Northcraft G.B., Lee C. & Lituchy T.R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal, 33,* 87-105.
- Dit proefschrift.

IV De beslissing om prestatie-informatie uit een ProMES prestatiesturingssysteem al dan niet te gebruiken voor beoordelings- en beloningsdoeleinden in een (hoofdzakelijk) individuele taaksituatie kan gekarakteriseerd worden door de uitspraak "Damned if you do, damned if you don't".

- Dit proefschrift.

V Deterministische modellen uit de regeltheorie/control theory hebben sterke beperkingen als metafoor voor menselijke motivatie, omdat zij geen plaats bieden aan een wezenlijk kenmerk van mensen, namelijk cognitieve processen.

- Locke E.A. & Latham G.P. (1990). *A theory of goal setting & task performance.* Englewood Cliffs, NJ: Prentice-Hall.

VI Onderzoek naar de invloed van feedback op arbeidsprestaties, waarin geen aandacht wordt geschonken aan de mediërende rol van doelen, getuigt ofwel van een gebrek aan theoretische kennis, ofwel van een bevooroordeelde toepassing van deze kennis. Overigens geldt hetzelfde voor onderzoek naar de invloed van goal setting op arbeidsprestaties, waarin geen aandacht wordt geschonken aan de modererende rol van feedback.

Bijvoorbeeld:
Kopelman R.E. (1982). Improving productivity through objective feedback: a review of the evidence. *National Productivity Review,* winter 1982-83, 43-55.
Alkemade N.D., Bissecker R.J. & Steensma H.O. (1994). Prestatieverbetering door objectieve feedback. *Gedrag en Organisatie, 7,* 65-70.

VII Het feit dat bij de rapportage van experimenteel onderzoek in wetenschappelijke tijdschriften de methodebeschrijving steeds vaker een ondergeschikte rol krijgt toebedeeld (beknopte samenvatting, klein lettertype) en mede daardoor vaak overgeslagen wordt bij lezing, kan tot gevolg hebben dat de conclusies van noodzakelijke nuanceringen worden beroofd.

- Latham G.P., Erez M., Locke E.A. (1988). Resolving scientific disputes by the joint design of crucial experiments by the antagonists: application to the Erez-Latham dispute regarding participation in goal setting. *Journal of Applied Psychology, 73,* 753-772.

VIII Als logisch gevolg van het voornemen om AIO's uitsluitend op basis van een studiebeurs promotie-onderzoek te laten doen, ligt het voor de hand om bij medewerkers van de vaste staf een korting toe te passen op het salaris naar rato van de door hen bestede tijd aan promotie-onderzoeksactiviteiten.

IX De stelling "meten is weten" is pas waar als aan het tegenovergestelde is voldaan.

X Het in één continue zitting uitspelen van schaakpartijen (in tegenstelling tot afbreken na 40 of 60 zetten), zoals voorgesteld door PCA-wereldkampioen Kasparov, zou van wedstrijdschaken weer een hoofdzakelijk individuele krachtmeting maken in plaats van een semi-geautomatiseerde teamsport.

XI Veel film-acteurs (m/v) stellen een 'intelligent' image op prijs. Hiertoe dragen zij in hun films en daarbuiten bij sommige gelegenheden steeds vaker een bril. Door vrijwel uitsluitend gebruik te maken van ongeslepen glas (sterkte 0), beledigen zij de intelligentie van de overige brildragers.

XII Gezien de voortdurende problemen bij de Nederlandse Spoorwegen zou het maandblad voor treinreizigers "Rails" beter omgedoopt kunnen worden tot "Op dood spoor".

XIII Meta-stellingen dienen verboden te worden.