Journal of Applied Psychology
2008, Vol. 93, No. 3, 540–567

0021-9010/08/$12.00 DOI: 10.1037/0021-9010.93.3.540

The Productivity Measurement and Enhancement System: A Meta-Analysis

Robert D. Pritchard, Melissa M. Harrell, Deborah DiazGranados, and Melissa J. Guzman
University of Central Florida

Meta-analytic procedures were used to examine data from 83 field studies of the Productivity Measurement and Enhancement System (ProMES). The article expands the evidence on effectiveness of the intervention, examines where it has been successful, and explores moderators related to its success. Four research questions were explored and results indicate that (a) ProMES results in large improvements in productivity; (b) these effects last over time, in some cases years; (c) the intervention results in productivity improvements in many different types of settings (i.e., type of organization, type of work, type of worker, country); and (d) moderator variables are related to the degree of productivity improvement. These moderator variables include how closely the study followed the original ProMES methodology, the quality of feedback given, whether changes were made in the feedback system, the degree of interdependence of the work group, and centralization of the organization. Implications based on these findings are discussed for future use of this intervention, and the system is discussed as an example for evidence-based management.

Keywords: productivity, productivity improvement, productivity measurement, performance measurement, Productivity Measurement and Enhancement System

Improving productivity in organizations is one of the cornerstones of industrial/organizational psychology and many tools have been developed to make these improvements. This article focuses on one intervention, the Productivity Measurement and Enhancement System (ProMES). ProMES is an intervention aimed at enhancing the productivity of work units within organizations through performance measurement and feedback. In this article, we use Pritchard's (1992) definition of *productivity,* that is, how effectively an organization uses its resources to achieve its goals.

The present meta-analysis describes studies conducted on the ProMES intervention and analyzes the results of an international collaboration lasting more than 20 years, in which the intervention was implemented in multiple settings by different researchers. The effectiveness of ProMES as an intervention has been previously described (e.g., Pritchard, 1995; Pritchard, Holling, Lammers, & Clark, 2002). In this article, we update the data on effectiveness of the intervention, examine effects over time, explore the effectiveness of the intervention in different settings, and focus on moderators that are related to the degree of this effectiveness. We first describe the theoretical background, then summarize how the intervention is done, show where it has been used, aggregate the results of these studies, and present the moderator findings.

Robert D. Pritchard, Melissa M. Harrell, Deborah DiazGranados, and Melissa J. Guzman, Department of Psychology, University of Central Florida.

We would like to thank all the many researchers who contributed to the ProMES database over the years. Winfred Arthur, Jr., Barbara Fritzsche, Gary Latham, Huy Le, and Carol Thornson made many helpful comments on earlier drafts of this article. The original funding for the Productivity Measurement and Enhancement System research effort was provided by the U.S. Air Force Human Resources Laboratory, which is now part of the Human Effectiveness Directorate of the U.S. Air Force Research Laboratory.

Correspondence concerning this article should be addressed to Robert D. Pritchard, Department of Psychology, University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390. E-mail: rpritcha@pegasus.cc.ucf.edu

Theoretical Background

The theoretical background of ProMES comes primarily from the motivational aspects of the Naylor, Pritchard, and Ilgen (1980; NPI) theory and a more recent motivation theory (Pritchard & Ashwood, in press) based on NPI theory. These theories are expectancy theories; they postulate that people are motivated by the anticipation of how their efforts will lead to satisfying their needs (e.g., Campbell & Pritchard, 1976; Heckhausen, 1991; Kanfer, 1990, 1992; Latham & Pinder, 2005; Mitchell & Daniels, 2003; Vroom, 1964).

The basics of the Pritchard–Ashwood theory are shown graphically on the left-hand side of Figure 1. The theory posits that people have a certain amount of energy, referred to as the energy pool, and that they have needs for such things as food, water, achievement, safety, and power. The energy pool varies across people and across time for any individual, and it is used to satisfy needs. The energy pool concept has similarities to the Kanfer and Ackerman (1989) and Kanfer, Ackerman, Murtha, Dugdale, and Nelson (1994) concept of attentional resources in that both deal with the issue of the limited resources people have for task performance. Motivation is the process that determines how this energy is used to satisfy needs. More specifically, the motivation process is defined as a resource allocation process through which energy is allocated across actions or tasks to maximize the person's anticipated need satisfaction.

The motivation process can be broken down into a series of components, shown on the right side of Figure 1. Energy is allocated across possible actions or tasks (e.g., a professor preparing lecture notes, writing manuscripts, or exercising). If energy is applied to actions, results are generally produced; typing (an

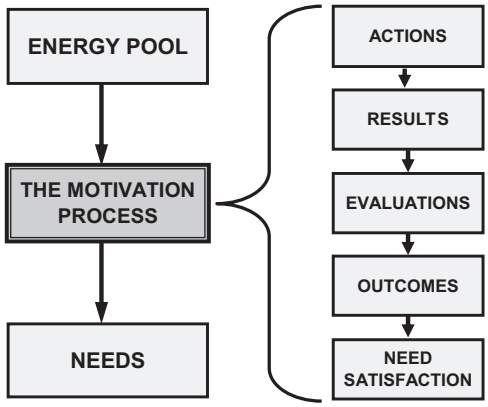

Figure 1. Pritchard and Ashwood (in press) motivation theory.

action) generates a manuscript (a result). Thus, a result is the person's output. When results are observed and an evaluator places the measured result on a good-to-bad continuum, this produces evaluations. Multiple evaluators evaluate the professor's manuscript, including the professor, colleagues who give feedback, journal reviewers, and readers of the eventual published manuscript. After these evaluations are made, outcomes occur. These are intrinsic outcomes, such as a feeling of accomplishment from writing a good paper, or extrinsic outcomes, such as forms of recognition, promotion, or pay raises for doing scholarly work. Outcomes derive their motivating power from their ties to need satisfaction. The more needs are satisfied, the greater the positive affect that is experienced; the less needs are satisfied, the greater the negative affect.

As with other expectancy theories, the linkages between the variables are critical. Between each of the boxes on the right side of Figure 1 are arrows that symbolize relationships called connections. The actions-to-results connection, shown by the arrow between the actions and the results, describes the person's perceived relationship between the amount of effort devoted to an action and the amount of the result that is expected to be produced. This perceived relationship can range from very strong to nonexistent.

The next arrow in Figure 1 refers to the results-to-evaluations connection. This connection reflects the person's perceived relationship between the amount of a result that is produced and the level of the evaluation that is expected to occur. There would be such a connection for each different result and for each person who evaluates the result(s), such as the professor, peers, supervisors, and researchers in other universities. The strength of these connections varies. The amount of scholarly outputs (a result) is probably strongly related to the department head's evaluation of the professor; amount of departmental service will likely have a much weaker relationship to overall evaluations. The evaluations-to-outcomes connection is the perceived relationship between the level of the evaluation and the level of outcome expected. The outcomes-to-need satisfaction connection defines the perceived relationship between how much of an outcome is received and the degree of anticipated need satisfaction that will result.

Although not shown in the figure, the result of these motivation components is the intent to behave. This leads to actual behavior, defined as the application of energy to actions. In turn, the actual behavior leads to actual results, evaluations, outcomes, and need satisfaction. These actual events have a feedback relationship with the components in the model. For example, actual outcomes received influence subsequent evaluations-to-outcomes connections.

All previous work-related expectancy theories include the idea of anticipated need satisfaction as well as relationships between effort and performance and between performance and outcomes (for reviews, see Latham, 2007; Latham & Pinder, 2005; Mitchell & Daniels, 2003). Our theory is different from other expectancy theories in several ways. First, it focuses on the resource allocation process rather than just on the overall level of effort. Second, it defines the connections between the variables as a type of nonlinear utility function, described below. Third, it considers the energy pool, an exhaustible amount of energy a person has at any point in time. Finally, this theory identifies the determinants of each connection (Pritchard & Ashwood, in press).

In addition to the two motivation theories, several other bodies of literature influenced the development of ProMES. These include the extant literature on feedback, goal setting, participation, roles and role ambiguity and conflict, and team effectiveness. How these literatures influenced the design of ProMES is described below.

The ProMES Intervention

We next describe ProMES and then show how it operationalizes the theory. The implementation of ProMES follows a series of steps. These are described most fully elsewhere (e.g., Pritchard, 1990; Pritchard, Paquin, DeCuir, McCormick, & Bly, 2002) and are summarized here. The first step is to form a design team composed of people from the target unit, the organizational unit that will ultimately use the measurement and feedback system. In addition to unit personnel, the design team includes one or two supervisors of that unit and a facilitator familiar with ProMES directing the design team. Design teams typically consist of 5–8 people; the total organizational unit typically contains between 5 and 35 members, occasionally as many as 50. The intervention results in a single set of objectives and quantitative indicators to be used for feedback. Organizational units larger than 50 are generally composed of subgroups with different objectives and indicators; thus, larger organizational units would usually have a ProMES system for each subunit.

Through discussion to consensus, this design team first develops the objectives of the department by identifying the tasks the department accomplishes for the broader organization. *Objectives* are typically general in nature, such as effectively dealing with production priorities, maximizing revenues, meeting training needs, optimizing customer satisfaction, and providing a safe working environment. Next, quantifiable measures called indicators are developed that assess how well these objectives are being met. *Indicators* are objective measures of output specific to that organizational unit. Examples include percentage of errors made, percentage of orders completed on time, number of clients seen, average time between failures of repaired items, and percentage of customers satisfied. Typically, there are 4–6 objectives and 8–12 indicators per organizational unit. These objectives and indicators are reviewed and approved by upper management to ensure that they are valid and aligned with broader organizational goals. Examples of actual objectives and indicators are shown in Table 1.

Table 1
Examples of Objectives and Indicators

Organizational Consultants
Setting: This unit worked with clients doing individual assessments of various types ranging from one-day assessment to multiple-day assessment centers.
Objective 1. Profitability
Indicator 1. Cost Recovery. Average amount invoiced per assessment divided by cost for that assessment.
Indicator 2. Billable Time. Percent monthly billable time on days when any assessment function is done.
Indicator 3. Billing Cycle Time. Average number of days between billing trigger and invoice submission.
Objective 2. Quality of Service
Indicator 4. Validity of Selection Assessments. Percentage of hits: people assessed predicted to be high performers who turn out to be high performers and those predicted to be marginal who are marginal. Index is based on a 6-month follow-up.
Indicator 5. Cycle Time. Percentage of assessment reports going out that went out on time.
Indicator 6. High Quality Experience of Participant. Percentage of participants giving "satisfied" and "very satisfied" ratings at the time of assessment.
Indicator 7. Customer Satisfaction. Percentage of "satisfied" and "very satisfied" to customer satisfaction measure.
Indicator 8. Consultant Qualifications. Percentage of licensable consultants who are licensed within 2 years of joining the firm.
Indicator 9. Ethics/Judgment Training. Percentage of staff with a minimum of 4 hours ethics/judgment training in the last 12 months.
Objective 3. Business Growth
Indicator 10. Assessment Revenue. Average revenues for the last 3 months from the assessment function.
Objective 4. Personnel Development and Satisfaction
Indicator 13. Personnel Skill Development. Number of actual tasks the person had been trained on divided by the number of possible tasks that person could be trained on.
Indicator 14. Personnel Satisfaction. Average number of "OK" and "good" days per person per month based on data entered when each person entered his or her weekly time card.
Photocopier Repair Personnel
Setting: Technicians go out on service calls to repair customers' photocopiers.
Objective 1. Quality: Repair and maintain photocopiers as effectively as possible.
Indicator 1. Mean copies made between service calls
Indicator 2. Percentage repeat calls
Indicator 3. Percentage of preventive maintenance procedures correctly followed
Objective 2. Cost: Repair and maintain photocopiers as efficiently as possible.
Indicator 4. Parts cost per service call
Indicator 5. Labor time per service call
Indicator 6. Percentage of repeat service calls caused by a lack of spare parts
Objective 3. Administration: Keep accurate records of repair and maintenance
Indicator 7. Percentage of required repair history information filled in correctly
Indicator 8. Percentage of parts warranty claims correctly submitted.
Objective 4. Attendance: Spend the available work time on work-related activities.
Indicator 9. Percentage of labor contract hours actually spent on the job.
Objective 5. Ambassadorship: Behave as correctly as possible on the job.
Indicator 10. Percentage of important social behaviors shown on the job as measured by customers' ratings.

The design team then develops contingencies. ProMES contingencies operationalize the result-to-evaluation connections in NPI theory and the Pritchard–Ashwood theory. A contingency is a type of graphic utility function relating the amount of each indicator measure to its value for the organization. An example contingency is shown in Figure 2. The indicator is the percentage of bed capacity used in the intensive care unit of a hospital. The horizontal axis shows levels of the indicator measure ranging from a low of 55% to a high of 85%. The vertical axis represents *effectiveness,* defined as the amount of contribution being made to the organization. The effectiveness scale ranges from −100, which is highly negative, through zero, which indicates meeting minimum expectations, to +100, which is highly positive. The function itself defines how each level of the indicator is related to effectiveness. A contingency is generated for each indicator. Once developed, contingencies are approved by upper management, as is done for objectives and indicators.

The design team develops contingencies using a discussion to consensus process such as that used with objectives and indicators. The basic idea is that the facilitator breaks down contingency development into a series of steps that the design team can implement. The first step is to identify the maximum and minimum realistic levels for each indicator. In the example bed capacity contingency, the design team decided that the minimum realistic value was 55% and the maximum realistic value was 85%. Next, the design team decides the minimum level of acceptable performance on each indicator. This is defined by the facilitator as the point of just meeting minimum expectations. Falling below this point would represent performing below minimum expectations on the indicator, and falling above this point would represent performing above minimum expectations. The design team, including the supervisor, discusses this value until consensus is reached. It is typical for the supervisor to bring management's point of view into this discussion. The point of minimum expectations then becomes

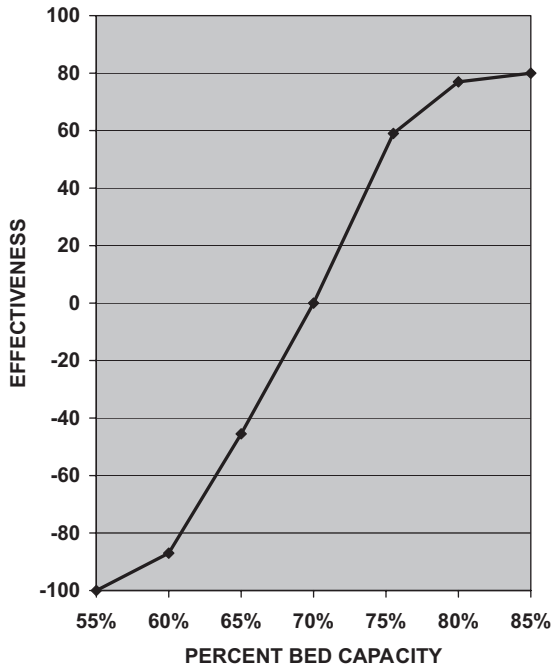

Figure 2. Example Productivity Measurement and Enhancement System (ProMES) contingency.

the point of zero effectiveness on the contingency. This value represents 70% bed capacity in the example contingency and, as shown in the graphic, is associated with a value of zero on the effectiveness scale. Next, the group ranks and then rates the effectiveness levels of the maximum and minimum indicator levels for each of the indicators. This process results in an effectiveness score for the maximum and minimum indicator levels for each contingency. The group then identifies, for each indicator, the general shape of the contingency between the zero effectiveness point and the maximum effectiveness value, and between the zero effectiveness point and the minimum effectiveness value, and then finalizes the location of the function in the contingency. As a final step, the group reviews the completed contingencies for accuracy. More detail on how contingencies are done can be found in Pritchard (1990).

Contingencies have several important features. The relative importance of each indicator is captured by the overall range in effectiveness score. Some indicators, such as the example in Figure 2, have large ranges. Others have smaller ranges. Those with larger ranges can contribute to or detract from the organization in greater amounts and are thus more important than those with smaller ranges. Contingencies also translate measurement into evaluation by identifying minimum expectations (the zero point) and how good or bad each level of the indicator is. For example, if the unit had a bed capacity of 75%, this would translate into an effectiveness score of +60, very positive and well above minimum expectations. Contingencies also identify nonlinearities, points where changes in indicator levels do not always result in the same amount of change in effectiveness. The figure shows a contingency with a point of diminishing returns, where the slope decreases with high bed capacity, indicating that further increases in bed capacity above 75% are not as valuable. Another feature of contingencies is that they offer a way of identifying priorities for improvement. One can readily calculate the gain in effectiveness that would occur if the unit improved on an indicator. For example, in the example indicator, going from a bed capacity of 70% to 75% would mean a gain in effectiveness of +60 points, but a gain from 75% to 80% would represent a gain of only +20 points. This suggests that improving bed capacity should be a high priority when it is below 75% but a lower priority when it is above 75%. Pritchard, Youngcourt, Philo, McMonagle, and David (2007) found evidence that units do use this improvement priority information that is provided by the contingencies. Finally, because contingencies rescale each measure to the common metric of effectiveness, a single overall effectiveness score can be formed by summing the effectiveness scores for each indicator. For example, if the effectiveness score for the bed capacity indicator was +60, it would be added to the effectiveness scores from the other indicators. If there were 12 indicators, 12 effectiveness scores would be summed to create the overall effectiveness score. This overall effectiveness score provides a single index of overall productivity.

Once the contingencies are approved by management, the feedback system is implemented. Unit personnel collect data on the indicators, and a printed feedback report is produced and distributed to each member of the target unit after each performance period. This feedback report includes a list of the objectives and indicators, the performance level on each indicator, the corresponding effectiveness score (e.g., for the example above, an effectiveness score of +60 for an indicator level of 75% bed capacity), and the overall effectiveness score, which is the sum of the effectiveness scores across the indicators. Also included are plots of indicator and effectiveness scores over time and a graphic presentation of the feedback. The length of the performance period varies across studies, but the most common period is 1 month.

A feedback meeting is held after each performance period to review the feedback report. As part of the feedback meeting, unit personnel identify ways of making improvements and use the feedback report to evaluate the success of improvement attempts made in the past. Individual units keep track of their own improvement efforts, as there is no formal mechanism in ProMES to record these. Feedback meetings continue over time in a continuous improvement model. The components of the measurement system, objectives, indicators, and contingencies are reviewed periodically to determine whether changes need to be made.

How ProMES Operationalizes the Theory

ProMES operationalizes key features of the motivation theory. Indicators are the operationalization of results. ProMES contingencies operationalize the results-to-evaluations connections. The action-to-results connections can be thought of as defining work strategies in that they identify how effort should be allocated across actions. Feedback reports and feedback meetings focus on developing better work strategies (i.e., a more optimal set of action-to-results connections). The feedback over time allows unit personnel to evaluate how well the new strategies are working and to refine them as needed.

Feedback. Scholars have argued for a number of important feedback features (Bobko & Colella, 1994; Erez & Kanfer, 1983;

Ilgen, Fisher & Taylor, 1979; London, 2003; Murphy & Cleveland, 1995; Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1988; Silverman & Wexley, 1984; Smither, London, & Reilly, 2005; Taylor, Fisher, & Ilgen, 1984; Wall & Lischeron, 1977; West & Anderson, 1996). The feedback system should include both a description and an evaluation of performance. This is done in ProMES by including both indicator and effectiveness scores. Because the system is known and totally transparent, people know what the evaluations will be. Validity of the measurement system in the sense of the measurement accurately reflecting the level of productivity, as well as perceived validity of the system, is maximized by carefully reviewing the indicators and contingencies in the design team, getting feedback from members of the unit not on the design team, and the management review. The high level of participation especially helps the perceived validity. This effort to ensure validity, maximize participation, make the system transparent, and give regular feedback helps with belief in the accuracy of the feedback. Reliability over time is maintained by using the same system over time. If changes need to be made, it is clear what has changed by the revision of indicators and/or contingencies. Agreement across evaluators occurs because the system is approved by the work unit, the supervisor, and upper management. The evaluators agree on the objectives, how success on these objectives will be measured, and how output on the measures translates into value to the organization (i.e., evaluations).

In their review of the effects of feedback on performance, Kluger and DeNisi (1996) found four feedback characteristics that were related to performance improvements across all of the analyses they performed. The largest effects from feedback occurred when the task was familiar, there were cues that supported learning, feedback provided information on discrepancies between performance and a standard, and the feedback did not threaten the self. ProMES is used with tasks that are well-known. The feedback reports and feedback meetings support learning new ways of doing the task. The effectiveness scores reflect deviation from the standard of minimum expected performance. The fact that the unit has participated in the design of the system and that feedback is typically done at the group level should reduce the threat to self.

Goal setting. ProMES also includes aspects of goal setting (Latham & Pinder, 2005; Locke & Latham, 2002). Whereas goal setting clearly includes formal, relatively public, agreed-upon levels of output to strive for (formal goal setting), it also includes less formal processes, such as private and public intentions to act (Frese & Zapf, 1994; Locke & Latham, 2002). ProMES implicitly, if not explicitly, includes many aspects of goal setting. First, ProMES provides feedback with regard to what employees need to start doing, stop doing, or continue doing to achieve a desired end state (i.e., performance goals). Feedback meetings focus on the behaviors necessary to attain desired end states; the benefits of focusing on behavioral goals have been discussed elsewhere (Brown & Latham, 2002; Latham, Mitchell & Dossett, 1978). Finally, and arguably most importantly, ProMES encourages the setting of learning goals, where people are urged to discover strategies or processes for attaining the desired outcome (Seijts & Latham, 2001).

Roles. Roles in NPI and Pritchard–Ashwood theories are defined as the set of results-to-evaluations connections. These connections identify expected outputs, indicate their relative value, and define how level of output is related to value and to evaluations. Role conflict and ambiguity have been linked to performance and attitude variables (Fisher & Gitelson, 1983; Jackson & Schuler, 1985; Tubre & Collins, 2000). Role ambiguity is reduced by specifically identifying the results-to-evaluations connections; role conflict is reduced by obtaining agreement on the connections by unit personnel, the supervisor(s), and management.

Participation. Whereas participation has shown conflicting findings (Wall & Lischeron, 1977; West & Anderson, 1996), there is considerable evidence that participation on issues of importance to employees can have positive effects on performance and attitudes, especially acceptance (Cawley, Keeping, & Levy, 1998; Dipboye & de Pontbriand, 1981; Locke & Schweiger, 1979). Participation is important, in part, because it enhances perceptions of procedural justice and voice (Cawley, Keeping, & Levy, 1998; Folger, 1977; Lind, Kanfer, & Earley, 1990; Lind & Tyler, 1988). Participation is a key part of ProMES. Most of the members of the design team are members of the unit, and these members are encouraged to discuss the development process with those not on the design team. In addition, the entire unit participates in the feedback meetings. These features should also increase perceptions of procedural justice and voice.

Teams. Literature on what makes teams effective has implications for ProMES. The intervention is primarily used with groups or teams and, when used with individuals, typically involves group feedback meetings with all of the individuals in the unit. In a major study of thousands of teams in the British National Health Service, West (2007) assessed three team characteristics: whether the team has clear objectives, whether members work closely together to achieve these objectives, and whether members meet regularly to review team effectiveness and how it could be improved. He found that the more these characteristics were present in the team, the greater the level of satisfaction and the lower the level of turnover intentions, errors, stress, injuries, and violence and harassment from patients and colleagues. ProMES includes all three of these characteristics.

Other research on teams reviewed by Salas, Rosen, Burke, Goodwin, and Fiore (2006) and by Salas, Kosarzycki, Tannenbaum, and Carnegie (2004) has identified characteristics that make teams more effective. These include holding shared mental models, having clear roles and responsibilities, engaging in a cycle of prebrief–performance–debrief, cooperating and coordinating, and using multiple performance criteria. Objectives, indicators, and contingencies can be seen as a type of shared mental model of the work that is developed by the group and then used in the feedback meetings. Roles and responsibilities are clarified through the measurement system and applied during feedback meetings. The ongoing feedback meetings are a type of prebrief–performance–debrief cycle where new ways of doing the work are developed and then evaluated in subsequent feedback meetings. Cooperation and coordination are encouraged through the feedback meetings. Multiple criteria of performance are included in the multiple indicators.

West (1990) hypothesized, and later empirically supported (Anderson & West, 1994; West, 1994), four dimensions of team climate that influence innovation: presence of a vision and shared objectives, participative safety, task orientation and climate for excellence, and group norms in support of innovation. The attempts to make improvements in task strategy made in the ProMES feedback meetings can be seen as a type of innovation.

ProMES fosters this climate for innovation. Shared objectives are the objectives and indicators; participative safety comes from the opportunity to participate in system design and in feedback meetings; task orientation is supported by the development of the measurement system, and task orientation and climate for excellence is supported by using the feedback to make improvements. Norms supporting innovation are fostered by using the feedback reports to make improvements. Agrell and Malm (2002) found that all four of these innovation climate dimensions improved after implementing ProMES in groups of Swedish police.

Another factor in team performance is *group reflexivity,* which is defined as "the extent to which group members overtly reflect on the group's objectives, strategies and processes, and adapt them to current or anticipated endogenous or environmental circumstances." (West, 1996, p. 559). The development of the measurement system and the feedback meetings are designed to promote group reflexivity. Agrell and Malm (2002) found that group reflexivity increased after the use of ProMES.

In conclusion, the theoretical foundation of ProMES comes from the motivational aspects of NPI and the Pritchard–Ashwood theories as well as literature on feedback, goal setting, roles, participation, and teams.

Research on ProMES

There has been considerable research on ProMES (Holling, Lammers, & Pritchard, 1998; Pritchard, 1990, 1995; Pritchard, Holling, et al., 2002; Pritchard et al., 1988; Pritchard, Jones, Roth, Stuebing, & Ekeberg, 1989; Pritchard, Kleinbeck, & Schmidt, 1993 [the German version of ProMES is called PPM: Partizipatives Produktivitaetsmanagement or Participative Productivity Management]; Pritchard, Watson, Kelly, & Paquin, 1998). Table 2 presents information from all available studies that meet the inclusion criteria described below for the present meta-analysis. The table shows the various types of organizations and jobs represented in the database. Studies have been conducted with personnel ranging from entry-level employees who did not finish high school to university faculty. Various types of organizations have been used, including military, profit and nonprofit, service, manufacturing, and sales organizations. Studies have been carried out in the United States, as well as in Australia, the Netherlands, Germany, Switzerland, Sweden, and Poland.

Research Questions

The present meta-analysis attempts to answer four research questions. Research Question 1 is whether ProMES is effective in improving productivity. There is a clear pattern of positive effects on productivity in the primary studies. The most recent summary of the ProMES literature, published in an edited book (Pritchard, Holling, et al., 2002), described results from 55 studies. The results reported in the current meta-analysis are based on 83 studies. This increased sample size allows for a more accurate estimate of the ProMES population effect size. Research Question 2 is the question of how long improvements last following the start of ProMES feedback. It is not uncommon for the effects of interventions to dissipate over time, and the question is, to what extent does this happen in ProMES interventions? Research Question 3 concerns how well the intervention works in different settings, such as various organization types, job types, and countries. In a recent meta-analysis, Parker et al. (2003) suggested that future organizational research should consider the role of setting and other moderator variables, such as those related to the organization's geographic location and size or the employee's level in the organizational hierarchy and his or her occupational group. Previous publications on ProMES did not include enough studies to make such comparisons possible. Research Question 4 concerned what factors influence the effectiveness of the intervention. This issue has not been addressed with meta-analytic techniques in previous publications on ProMES.

These research questions are addressed with the data from the available ProMES projects, described below. The primary criterion variable used to assess the effectiveness of ProMES was change in performance from baseline, prior to the start of ProMES feedback, to performance once ProMES feedback begins. Each project has an overall effectiveness score for each period of baseline and feedback. This overall effectiveness score is the sum of the effectiveness scores over all the indicators for a given feedback period and thus serves as the best measure of overall performance. Research Questions 1 and 2 were addressed with these data. Research Question 3 was addressed through moderator analyses with variables such as country and organization type. The question addressed was whether the amount of performance improvement after ProMES feedback was different for these different subgroups of projects. Research Question 4 was also addressed through moderator analyses with variables such as amount of prior feedback, number of personnel in the target unit, and degree of management support.

Moderators and the Meta-Analysis Questionnaire

Early in the ProMES research program, it became clear that many studies were being conducted by different research teams. In anticipation of a future meta-analysis, such as the one reported here, a questionnaire was developed (Paquin, 1997). This questionnaire attempted to identify variables that might influence the effectiveness of the intervention. To develop this questionnaire, the literature on the effectiveness of interventions, such as ProMES, was reviewed and a qualitative analysis was done on the basis of interviews conducted with researchers in multiple countries who were using the intervention. A draft version was then prepared and reviewed by ProMES researchers in the various countries for clarity, ease of use, and comprehensiveness. These comments were used to develop the final instrument.

This final questionnaire contains five major sections. The section labeled Overall Project Description includes an overall description and purpose of the project. The second section, Characteristics of the Organization, measures size, type, centralization, degree of trust between unit personnel and management, initial attitudes toward ProMES, and support by management. The third section, a Description of the Developed System, includes characteristics of the process, such as the composition of the design team, how design team meetings were conducted, characteristics of the final system such as number of objectives and indicators, how the feedback was delivered, and how feedback meetings were conducted. The fourth section, Reactions to the System, includes favorableness of reactions by unit personnel, supervisors, management, and unions; features that these groups liked and disliked; and

Table 2
Studies in the Productivity and Enhancement Management System (ProMES) Database

Setting	Country	*d*	*N*		Source
			Baseline	Feedback	
USAF: Electronic repair	United States	3.27	9	25	Pritchard et al., 1988, 1989
USAF: Warehouse: Receiving	United States	4.83	8	20	Pritchard et al. 1988, 1989
USAF: Warehouse: Storage and issue	United States	5.31	8	20	Pritchard et al., 1988, 1989
USAF: Warehouse: Inspection	United States	3.32	8	20	Pritchard et al., 1988, 1989
USAF: Warehouse: Pickup and delivery	United States	3.20	8	20	Pritchard et al., 1988, 1989
Rubber factory: Top management	Germany	1.83	12	3	Kleinbeck & Fuhrmann, personal communication, 2000
Rubber factory: Production	Germany	1.58	12	3	Kleinbeck & Fuhrmann, personal communication, 2000
Pension fund	Germany	2.32	8	10	Werthebach & Schmidt, personal communication, 2001
Assembly center	Germany	1.40	12	12	Przygodda et al., 1995
Coiling center	Germany	2.68	10	10	Przygodda et al., 1995
Industrial/organizational psychology program	United States	4.61	3	5	Jones et al., 2002
Electrical equipment manufacture: 1	United States	0.11	1	4	Jones, personal communication, 1998
Electrical equipment manufacture: 2	United States	−0.66	1	2	Jones, personal communication, 1998
Electrical equipment manufacture: 3	United States	−0.37	1	8	Jones, personal communication, 1998
Electrical equipment manufacture: 6	United States	−0.52	1	4	Jones, personal communication, 1998
Electrical equipment manufacture: 8	United States	−0.80	1	4	Jones, personal communication, 1998
Electrical equipment manufacture: 9	United States	−2.50	1	4	Jones, personal communication, 1998
Electrical equipment manufacture: 10	United States	0.47	1	4	Jones, personal communication, 1998
Electrical equipment manufacture: 12	United States	0.10	10	11	Jones, personal communication, 1998
Photocopier service: 1	Netherlands	2.15	1	25	Kleingeld et al., 2004
Photocopier service: 2	Netherlands	2.15	1	25	Kleingeld et al., 2004
Swiss school	Switzerland	4.71	3	1	Minelli et al., 2002
Social Security clerical: 1	Sweden	0.09	2	10	Malm, personal communication, 1997
Social Security clerical: 2	Sweden	−0.32	2	10	Malm, personal communication, 1997
Oil company clerical staff	Netherlands	0.24	4	1	Miedema et al., 1995
Sports equipment manufacturing	United States	0.71	13	7	Jones, 1995a
Chemical manufacturing	United States	1.91	9	5	Jones, 1995b
Battery manufacturing	United States	−0.55	22	8	Jones, 1995b
Cardboard manufacturing shipping	Netherlands	1.24	12	24	Janssen, personal communication, 1996
Cardboard manufacturing: 1	Netherlands	2.58	1	7	Janssen et al., 1995
Cardboard manufacturing: 2	Netherlands	2.04	12	12	Janssen, personal communication, 1996
Cardboard manufacturing: 3	Netherlands	0.80	12	12	Janssen, personal communication, 1996
Cardboard manufacturing: 4	Netherlands	3.11	4	12	Janssen, personal communication, 1996
Cardboard manufacturing: 5	Netherlands	−1.11	4	6	Janssen, personal communication, 1996
Cardboard manufacturing: 6	Netherlands	1.39	12	3	Janssen, personal communication, 1996
Cardboard manufacturing: 7	Netherlands	2.09	9	3	Janssen, personal communication, 1996
Computer maintenance and repair: 1	Australia	1.04	19	11	Bonic, 1995
Computer maintenance and repair: 1	Australia	2.46	14	5	Bonic, 1995
Photocopier service	United States	0.21	15	10	Howell, et al., 1995
Poultry processing: 1	Netherlands	0.51	1	16	Van Tuijl, personal communication, 2001
Poultry processing: 2	Netherlands	0.92	1	16	Van Tuijl, personal communication, 2001
Confectionary plant	Hungary	1.82	20	9	Van Tuijl, personal communication, 2001
Traffic police Group 1	Sweden	2.78	6	59	Agrell & Malm, 2002
Traffic police Group 2	Sweden	4.19	6	59	Agrell & Malm, 2002
Traffic police Group 3	Sweden	3.22	6	59	Agrell & Malm, 2002
Traffic police Group 4	Sweden	1.88	1	26	Agrell & Malm, 2002
Government trainers	United States	2.60	4	1	Clark, 1999
Life insurance sales Group 1	Switzerland	0.96	4	12	Minelli, 2005
Life insurance sales Group 2	Switzerland	1.99	4	12	Minelli, 2005
University research support: Clerical 1	United States	0.18	1	38	Pritchard, personal communication, 2004
University research support: Clerical 2	United States	0.97	7	30	Pritchard, personal communication, 2004
University research support: Clerical 3	United States	−0.03	1	29	Pritchard, personal communication, 2004
University research support: Clerical 4	United States	0.02	1	27	Pritchard, personal communication, 2004
University research support: Clerical 5	United States	0.41	10	22	Pritchard, personal communication, 2004
University research support: Clerical 6	United States	−0.14	18	12	Pritchard, personal communication, 2004
University research support: Human resources	United States	−0.95	5	21	Pritchard, personal communication, 2004
University research support: Treasury	United States	0.24	5	22	Pritchard, personal communication, 2004
University research support: Management information systems	United States	0.30	6	11	Pritchard, personal communication, 2004

Table 2 (*continued*)

Setting	Country	*d*	*N*		Source
			Baseline	Feedback	
University research support: Clerical 7	United States	2.22	5	18	Pritchard, personal communication, 2004
University research support: Clerical 8	United States	0.00	5	15	Pritchard, personal communication, 2004
University research support: Clerical 9	United States	−0.46	3	13	Pritchard, personal communication, 2004
University research support: President's office	United States	1.98	1	11	Pritchard, personal communication, 2004
University research support: Senior management	United States	0.51	1	8	Pritchard, personal communication, 2004
Internal Revenue Service: 1	Sweden	−0.54	1	28	Malm, personal communication, 2002
Internal Revenue Service: 1	Sweden	−0.96	1	24	Malm, personal communication, 2002
Military equipment agency	Sweden	−0.32	1	3	Malm, personal communication, 2004
Air traffic controllers: 1	Sweden	2.20	1	4	Malm, personal communication, 2002
Air traffic controllers: 2	Sweden	2.57	1	7	Malm, personal communication, 2002
Marketing research	Germany	1.00	2	2	Roth & Moser, 2005
Microscope assembly: 1	Germany	1.13	1	14	Hollmann, personal communication, 2004
Microscope assembly: 2	Germany	1.38	4	11	Hollmann, personal communication, 2004
Microscope assembly: 3	Germany	0.97	2	11	Hollmann, personal communication, 2004
Psychiatric hospital: Top management	Germany	1.23	1	2	Hollmann, personal communication, 2006
Psychiatric hospital: Nurses 1	Germany	0.00	1	40	Hollmann, personal communication, 2006
Psychiatric hospital: Nurses 2	Germany	0.48	1	27	Hollmann, personal communication, 2006
Psychiatric hospital: Technicians	Germany	−1.20	1	18	Hollmann, personal communication, 2006
Psychiatric hospital: Technicians	Germany	1.37	1	5	Hollmann, personal communication, 2006
Psychiatric hospital: Technicians	Germany	2.81	1	41	Hollmann, personal communication, 2006
Manufacturing: Production planning	Germany	1.60	1	11	Hoschke, personal communication, 2000
Manufacturing: Cutting	Germany	0.43	1	8	Hoschke, personal communication, 2000
Manufacturing: Assembly	Germany	0.17	1	8	Hoschke, personal communication, 2000
Manufacturing: Maintenance	Germany	−0.21	1	10	Hoschke, personal communication, 2000
Manufacturing: Managers	Germany	−1.39	1	8	Hoschke, personal communication, 2000

Note. *d* = effect size; *N*s represent the number of baseline and feedback periods, respectively. USAF = U.S. Air Force. Sources with dates are published and cited in the References with asterisks. Personal communication represents either an unpublished study or information provided by the project director.

changes made to the system over time. The final section, Project Data, includes the objectives, indicators, contingencies, and effectiveness scores over time.

The response scales differed depending on the item content. Some were open-ended, some asked the respondent to indicate percentages for different anchors, and some were checklists, but most were 5-point Likert scales. The specific items used in this article are presented below.[1]

Researchers conducting ProMES projects were asked to complete this questionnaire about their study. Thus, the meta-analysis data collection was different from that done in typical meta-analysis studies. In the vast majority of our cases, the data came from the researcher conducting the study and not from a third party coder, working from a published or unpublished description of the study. Our approach has advantages and disadvantages.

The major advantage is that the researcher knows much more about the study and can provide more information than a rater can obtain from a description written by someone else. This allows for richer detail than is normally possible in a meta-analysis. Another advantage is that data for the meta-analysis can be collected on studies that are not written in report or publication form, thus increasing the number of studies for the analysis.

The major disadvantage is that the people completing the questionnaire cannot be trained and assessed for interrater reliability in the way coders are trained and evaluated in a typical meta-analysis. Specifically, researchers cannot be given a sample of studies conducted by a third party and then asked to code them for comparison with other coders' ratings of the same studies. Nor is it possible to obtain such reliability estimates after the questionnaires are completed because the majority of the information comes from the researcher's knowledge of that specific project, not from a written document such as a publication. For example, a publication would rarely contain information on variables such as amount of prior feedback, trust between unit personnel and management, turnover, level of management support, or stability of the organization's external environment. Even if a subset of the articles in a typical meta-analysis would include data on these variables, this information would not be included in every study nor would the definitions of these variables be the same across studies.

Selection of Moderator Variables

One major challenge in this research was the selection of moderator variables for examination. There are over 100 possible variables assessed by the questionnaire that could have been used as moderators. The total number of studies included in the database was 83; however, complete data were not available for every questionnaire variable. Clearly, all of the questionnaire variables could not be included in the analysis because of power considerations given the number of studies.

[1] The full measure can be found at http://promes.cos.ucf.edu/meta-structural.php

Several approaches were used to deal with this dilemma. The main criterion involved selecting variables with research- and theory-based rationales. The specific rationales for each of the variables selected are presented below. The overall variable selection strategy started with the identification of variables dealing with how well the feedback was performed. This was done because feedback is such a central feature of ProMES. Next, we wanted to include variables at multiple levels of analysis. Most ProMES studies do not have individual-level data, but both work group and organization-level data are available. We selected variables that represented both of these levels of analysis.

To reduce the number of variables, composite scales of related items were formed on the basis of conceptual similarity. For example, the quality of feedback composite was composed of five items from the meta-analysis questionnaire: percentage of feedback reports followed by a feedback discussion meeting, percentage of feedback meetings with supervisor present, time taken for typical feedback meeting, percentage of feedback meeting time focused on constructive behaviors initially, and percentage of time focused on constructive behaviors after experience with the system. Next, coefficient alpha was calculated for each scale, and items were added or removed as necessary to maximize coefficient alpha while maintaining a theoretically sound scale. The specific items making up each composite, response scales, and internal consistency reliabilities are presented in the Method section.

Feedback System Moderators

Although ProMES is a complex intervention, the key process is feedback. There is a substantial amount of evidence indicating that feedback generally has a positive impact on performance. However, the strength and nature of this relationship have varied considerably across studies (Kluger & DeNisi, 1996; Kopelman, 1986). In fact, Kluger and DeNisi found negative relationships between feedback and performance in over one third of the studies included in their meta-analysis; thus, it is clear that the presence of feedback is not always sufficient to improve outcomes.

Quality of feedback. Kluger and DeNisi (1996) make it clear that the way feedback is provided has a major impact on its effectiveness. For example, evaluation standards should be clear, descriptive, specific, and developed with the help of those to whom the standards apply (Bobko & Colella, 1994). Likewise, information resulting from such evaluations should be specific, provided regularly, and stated descriptively in behavioral terms (Balcazar, Hopkins, & Suarez, 1986; Kopelman, 1986; Taylor, Fisher, & Ilgen, 1984).

Thus, feedback is a critical process in ProMES. To develop an index of feedback quality, we selected several feedback characteristics to develop into a feedback quality composite and predicted that feedback quality should be related to the effectiveness of the intervention. The actual items for this and the other scales are presented in the Method section.

Prior feedback. The amount of feedback prior to the ProMES intervention could influence its effectiveness. As Ilgen et al. (1979) noted, feedback should have information value to the recipient. This means that the feedback should provide an incremental increase in information about behavior over and above what is already known by the individual. An organizational unit where the level of prior feedback is high was predicted to show smaller productivity gains with the additional feedback from ProMES compared with a unit where level of prior feedback was low.

Changes in the feedback system. A final variable that could influence the effectiveness of the feedback system is the number of changes that had to be made to the ProMES system once it was instituted. As noted above, ProMES can result in a specification of roles, thus reducing role ambiguity and role conflict. Reducing ambiguity and conflict should have positive effects on performance (Fisher & Gitelson, 1983; Jackson & Schuler, 1985; Tubre & Collins, 2000). However, a high number of changes could indicate confusion about the system, a system that was not well designed to start with, or changes in the nature of the work, such as the introduction of new equipment that required such changes. These changes could increase role ambiguity and conflict and, in turn, decrease the effectiveness of the feedback system.

Degree of Match With Original Intervention Methodology

The variable termed degree of match represents the degree to which the steps done in the intervention matched those described in the Pritchard (1990) book. As is common in field studies, the researchers did not have total control over the process, and the researchers in some studies could not follow all of the steps originally specified. For example, contingencies were developed by supervisors rather than by the people in the work unit, or feedback meetings were not held in a location conducive for discussion of the feedback report. The issue concerned the extent to which deviating from the original process reduced the productivity improvement. The expectation was that such changes would decrease the effects of the intervention on performance (Bobko & Colella, 1994; Erez & Kanfer, 1983; Murphy & Cleveland, 1995; West & Anderson, 1996).

Work Unit–Level Moderators

Trust. *Trust* has been defined as the willingness of an individual or a group to be vulnerable to the actions of another individual or group on the basis of the expectation that the other will perform a particular action important to the trustor, irrespective of the trustor's ability to monitor or control that other (Mayer, Davis, & Schoorman, 1995). We focused on the trust between work unit personnel and their management, as suggested by Ergeneli, Ari, and Metin (2007). Although previous research on trust has been unclear regarding what factors contribute to trust (Cook & Wall, 1980; Kramer, 1999), we predicted that the level of trust between management and lower level employees prior to the intervention would influence the productivity gains. Specifically, where trust is high, managers may give employees more control and be more willing to approve the measurement systems developed by design teams. It could also be argued that settings where trust between management and unit personnel is low could stand to gain more from an intervention, such as ProMES, because agreeing on the components of the system could increase trust, and this gain could have a positive effect on performance (Sargent & Pritchard, 2005). Thus, there are arguments in favor of both a positive and a negative relationship; this is reflected by the results of a recent meta-analysis demonstrating considerable heterogeneity in the relationship between trust and organizational outcomes across studies

(Dirks & Ferrin, 2002). Therefore, no directional predictions were made for trust.

Number of personnel in target unit. Social impact theory (Latane, 1981; Latane & Nida, 1980), social loafing (Karau & Williams, 1993; Latane, Williams, & Harkins, 1979), and diffusion of responsibility (Darley & Latane, 1968) all posit that as group size increases, performance on a wide variety of activities decreases. Thus, we expected that the larger the ProMES target unit, the smaller would be the gain in productivity.

Type of worker. Different types of workers, such as blue-collar, managerial, and academic, have very different types of jobs with varying demands, control, and expectations for feedback (Frenkel, Tam, Korczynski, & Shire, 1998; Korczynski, 2001; Taylor, Mulvey, Hyman, & Bain, 2002). Thus, it seems reasonable that different types of workers could respond differently to interventions such as ProMES. Paquin (1997) found subgroup differences and large variances when examining ProMES intervention effectiveness across types of workers.

Turnover. Turnover was defined as the percentage of personnel who left the organization annually during the project. As with all of the other moderator variables, actual items are shown in the Method section. One possible result of turnover is the disruption of performance when experienced employees leave and new employees join the unit (Abbasi & Hollman, 2000; Beadles, Lowery, Petty, & Ezell, 2000; McElroy, Morrow, & Rude, 2001; Mobley, 1982; Price, 1989). The effects of ProMES may also be decreased by turnover because new employees did not participate in the design of the feedback system and thus did not understand it. Therefore, employees who did not participate in, accept, or understand the system may leave. Watrous, Huffman, and Pritchard (2006) found some evidence that degree of management turnover was negatively related to performance gain under ProMES. Thus, although there are arguments in favor of both a positive and a negative relationship, we predicted that the higher the turnover, the lower would be the effectiveness of the intervention.

Complexity of the work unit. The level of complexity of the work, including complexity of the technology, structure complexity, and complexity of the demands on the unit, could influence ProMES effectiveness. Murphy's (1989) model of performance suggests that job complexity plays a role in the nature of job performance, and other research has shown that job complexity moderates a number of relationships with job performance (Keil & Cortina, 2001; Sturman, 2003). Feedback could be more important when complexity is high because feedback such as that in ProMES, which is clear and accurate and includes all of the important parts of the work, may reduce the cognitive load of processing performance information on multiple, interrelated tasks. Lindell (1976) argued that cognitively oriented feedback can have the effect of disentangling cue–criterion relationships and eliminating nonessential noise and that both of these effects of feedback become more important as complexity increases. Youmans and Stone (2005) suggested that as judgment tasks become more complex, feedback and task information would have a positive effect on performance. On the basis of these arguments, we predicted complexity to be positively related to productivity gain.

Interdependence. Interdependence is the degree to which group members require interaction for achieving results (Saavedra, Earley, & van Dyne, 1993). Higher levels of group interdependence signal a greater need for teamwork (Smith-Jentsch, Johnston, & Payne, 1998) and a greater need to evaluate the team's processes along with its outputs (Balzer, Doherty, & O'Connor, 1989; Brehmer, 1980; Garafano, Kendall, & Pritchard, 2005). Because ProMES feedback is largely output feedback rather than process feedback, it may be less effective in more interdependent groups. Although the feedback meetings in ProMES are designed specifically to address process issues in how the task is done, we would still expect that the greater the interdependence, the greater the need for process feedback. Thus, ProMES feedback should be less effective for units with high levels of interdependence than for units with less interdependence.

Management support. Klein and Sorra (1996) suggested that a major factor in the success of an intervention is supportive management. Rodgers and Hunter (1991) conducted a meta-analysis, which showed that management support moderated the effects of interventions involving management by objectives. Specifically, intervention effects on performance were positively related to level of management support. Thus, we predict that the higher the management support for a ProMES intervention, the greater will be its subsequent effectiveness.

Organization-Level Moderators

Centralization. Centralization is the degree to which decision-making authority is held by a few high-level managers as opposed to being dispersed throughout the organization (Dalton, Todor, Spendolini, Fielding, & Porter, 1980). It is seen as related to autonomy in that autonomy is at one end of the continuum and centralization is at the other (Massie, 1965; Sisk & Williams, 1981). If an organization is highly centralized, employees have less authority to make decisions in the organization. Therefore, because of the gain in autonomy, implementing a participative intervention system such as ProMES is hypothesized to yield greater productivity gain in more centralized organizations.

Stability of the organization's external environment. Stability of the external environment is the degree to which external demands, requirements, and external stakeholders remain consistent over time (Venkatesh & Speier, 2000). The stability of an organization's environment can contribute to the social context for productivity improvement because management is more externally focused if the environment is unstable and thus less focused on internal processes (David, 2005). Miedema, Thierry, and van Oostveen (1995) described stability in the organization's environment as an important factor when considering the implementation of ProMES. Thus, the more stable the organization's environment, the greater the productivity increases from an intervention such as ProMES.

Function of the organization and type of organization. It was also possible that organizations with different functions (e.g., manufacturing, service, sales) or of different types (e.g., private for profit, not for profit) would show different effects of the intervention. For example, Guzzo, Jette, and Katzell (1985) found that interventions conducted in government organizations resulted in larger effect sizes than those conducted in for-profit or nonprofit organizations. We made no specific predictions here about which functions or types would show the largest effects; the question was whether the intervention is effective in different types of organizations.

Country. There may be differences in the effectiveness of ProMES due to country. Cross-cultural studies (e.g., Hofstede, 1980; House, Hanges, Javidan, Dorfman, & Gupta, 2004) have shown country differences in such variables as collectivism, power distance, doing orientation, and determinism. It seems logical that such differences could influence the success of interventions. Rafferty and Tapsell (2001) as well as Kirkman and Shapiro (2001) argued that differences in the effectiveness of self-managed work teams can be attributed to cultural influences. As with the type of organization variable, we made no specific predictions about which countries would have the largest effects; rather we searched to determine whether ProMES is effective in different cultural contexts.

Method

Data Set and Inclusion Criteria

The entire ProMES database includes all of the studies, published or not, where the data were made available to us. ProMES researchers known to us were contacted throughout the past 17 years to provide ongoing information on any published or unpublished studies. A search of the published literature was done in PsycINFO, Business Source Premier, ProQuest Dissertations and Theses, and Google Scholar. The search terms were ProMES, Productivity Measurement and Enhancement System, and PPM (Partizipatives Produktivitätsmanagement [Participative Productivity Management], the German term for ProMES). This search yielded 16 empirical publications, many with multiple ProMES studies. All of these published studies were known to us and were already part of the database of 88 studies discussed below. Table 2 shows the studies from each publication as well as the other, unpublished studies. Performance data were made available from all of these studies, although complete questionnaire data on the moderators were not always provided.

We estimate that there were 3 to 5 studies completed that never became part of the ProMES database because they were done some time ago and the data were not available. There was one set of 4 to 5 more recent studies where the data were not provided to the researchers. Five to 8 studies of which we were aware were discontinued as a result of organizational changes before ProMES feedback could be started; therefore, the complete intervention was not implemented, and performance data during feedback were not available. Thus, the resulting database used here represents approximately 90% of the known ProMES studies both published and unpublished that were completed since 1987, the publication date of the first ProMES study.

Each ProMES study includes a baseline period where data are collected on the indicator measures but no feedback is given. This is followed by a feedback period where unit personnel receive feedback after each work period and meet to discuss these feedback reports. A study had to have at least three periods of combined baseline and feedback periods. This criterion was necessary to be able to calculate the effect sizes used here, which are discussed below. Five studies from the full database of 88 studies failed to meet this criterion.

The Database

Table 2 lists information on the 83 studies used in the analyses. The combined number of baseline and feedback periods for these studies ranged from 3 periods to 65 periods, with a mean of 19.84 periods. The number of baseline periods ranged from 1 to 22, with a mean of 5.23. Number of feedback periods ranged from 1 to 59, with a mean of 14.67. The time period used for feedback varied across studies and depended on the work cycle time and how frequently indicator data were available. In most studies, feedback was given each month. Of the 71 studies for which this datum is available, 56 (78.87%) used monthly feedback, 9 (12.67%) used more frequent feedback, and 6 (8.46%) used less frequent feedback.

Dependent Variable

The primary dependent variable in a ProMES study is productivity improvement. Productivity is operationalized by the overall effectiveness score, the total of the effectiveness scores from each of the indicator measures. The primary evaluation criterion for a ProMES project, productivity improvement, is the standardized change in this overall effectiveness score from baseline to feedback.

Effect Sizes

The overall effectiveness score was used to calculate effect sizes with the *d* statistic (Hunter, Schmidt, & Jackson, 1982). To calculate the effect size, we computed the mean difference in the overall effectiveness scores between the feedback and baseline periods. This difference was then divided by the pooled standard deviation. The pooled standard deviation was calculated by first summing the squared deviations of the baseline overall effectiveness scores for each period from the baseline mean. This sum was added to the sum of the squared deviations of the feedback overall effectiveness scores for each period around the feedback mean. The total sum of squares was then divided by the combined degrees of freedom, and the square root of the result was the pooled standard deviation.

Calculating these effect sizes requires that there be at least three periods of data, with at least one period at baseline and one in feedback. Data from both baseline and feedback stages are necessary to calculate the mean difference; two periods of baseline or feedback data are required so that the pooled standard deviation can be calculated. As noted above, five studies did not meet these criteria and were excluded from the analyses.

This approach to calculating an effect size is different from that used in most meta-analyses. Typically, there is a set of scores from one time period (e.g., prior to the intervention) and a second set of scores after the intervention. Each pair of scores comes from one individual or group, and the scores for each pair are from the same measure. The pooled standard deviation is calculated using the sum of squares around the mean for the first time period and adding this to the sum of squares from the second time period. In contrast, our data have a single score for each case at multiple points in time during baseline and multiple points of time during feedback.

The more typical effect size was not appropriate here. It would be possible to calculate the mean overall effectiveness score for baseline and the mean for feedback and then use each study as a pair of observations. A single effect size could then be calculated across all studies. This approach was not appropriate for our

purposes because the overall effectiveness score in one study is not comparable to this score in another study. Studies with different numbers of indicators and different effectiveness score ranges in the contingencies will have very different maximum overall effectiveness scores. For example, a study where there were 12 indicators will likely have larger overall effectiveness scores for the same relative performance than a study with 8 indicators because the overall effectiveness score is a sum of indicator effectiveness scores. Furthermore, this procedure would produce a single effect size across all studies and thus preclude tests for moderators.

Another approach would be to calculate effect sizes at the level of the indicators within one study. The average effectiveness score for each indicator during baseline and the average during feedback could be used to calculate an effect size across the indicators. This method would produce a very small sample size for each effect size: the number of indicators used in that study. Use of this approach would also reduce the sample size substantially because complete indicator data are required, and only 49 of the 83 studies have complete indicator data. Even ignoring these considerations, using this approach still would not be comparable to typical derivation of effect size because the pre–post scores of each pair were not taken on the same measure, and the measures did not use the same scale, as is the case for typical effect sizes; thus the pooled standard deviation would be inflated.

Another issue with the current approach used to calculate effect sizes is that, in essence, it assumes the independence of the overall effectiveness scores in the different time periods for a given unit. This was clearly not the case in the current study because the effectiveness scores come from the same people at different points in time. This would suggest the need for an index that captures this dependency. However, it is not clear how such dependency would be calculated in this setting. One could argue that such dependence would lead to inflated effect sizes because the variability across the scores would be reduced given that they come from the same people. One could also argue that dependence could result in lower effect sizes because removing any common variance due to the dependency would decrease the variability estimate, thereby increasing the effect size. In one sense, it is not a critical issue. It is clear from the results described below that the effect sizes are very large, and regardless of how they are calculated, the conclusion of large effects would be unchanged.

There may be limitations to the type of effect size calculation used here, and it may not be appropriate to compare our effect sizes with results from more typical effect sizes. However, we believe this is the best type of effect size for these data and that it is appropriate to report them and use them for moderator analyses.

Comparison Groups and Attitude Measures

Some studies had comparison groups, that is, units doing similar work which did not receive the ProMES intervention. Data were collected on key output measures for these groups. A number of studies also measured attitudes, such as job satisfaction, turnover intentions, stress, and aspects of organizational climate. As summarized in Pritchard, Paquin, et al. (2002), most studies showed an improvement on these measures after feedback, and none showed a decrease. However, there were not enough studies measuring the same attitudes to be usable for this meta-analysis.

Independence

One issue in meta-analysis is the independence of the studies. Results are considered dependent when multiple measures of the same dependent variables are collected from the same sample of participants (e.g., Arthur, Bennett, Edens & Bell, 2003). By this criterion, there was no dependence in this database. A single group of participants, in our case one work unit, provided data for only one effect size. However, as discussed above, there was dependency in the data used to estimate an effect size in each primary study because an effect size is based on performance measures of a single group at different periods of time. However, this type of dependency is an essential part of this type of measure. Finally, there were some cases where a single researcher or facilitator conducted multiple studies that were part of the database. In some of these cases, there were multiple work groups in a given organization. In other cases, the same person conducted multiple studies in different organizations. In most of these cases, however, the actual facilitator was different for the different studies. For example, Robert D. Pritchard is responsible for 19 of the studies, but 7 different facilitators worked with these groups.

Moderator Variables

The moderators that are continuous variables are shown in Table 3. This table shows the variable, the questionnaire item(s) used to measure each, and the extremes of the scale anchors. If the variable is a composite, the internal consistency reliability is shown. Composites were formed by averaging the component items. In cases where the response scales for the items making up a composite differed, items were standardized before calculating the mean.

One research question was whether ProMES is effective in different settings. The different settings are captured by the categorical variables of country, type of organization, function of the organization, and type of worker. The country variable was the country where the project was done. The other categorical variables were measured by items giving a list of response options. The type of organization variable came from a list of three possible types: private for profit; private nonprofit; and government/public. Function included a list of manufacturing, sales, service, educational, research, military, and health care. Type of worker included a list of managerial/professional, blue-collar/labor, technician, sales, clerical/office, academic/teaching, and other. Some grouping was done for some variables to provide a sufficient number of studies. For example, function of the organization was grouped into manufacturing, sales, and several other functions that were all service oriented in nature (service, education, health care, and military). Four countries included a sufficient number of studies. Germany and Switzerland were combined because the Swiss studies were conducted in the German-speaking part of Switzerland. The specific categorical moderator variables are shown in Table 4.

Results

Overall Effects on Productivity

The first research question was whether the ProMES intervention improved productivity. An overall picture is shown in Figure 3, where the mean overall effectiveness score over baseline and feedback time periods is plotted. Overall effectiveness is the sum

Table 3
Continuous Variables

Variable	Item(s)
Degree of match	Overall, how closely did the development and implementation of the system in this setting match the process outlined in the 1990 ProMES book? 5. Very closely: That process was followed as closely as possible. 3. Moderately: A few meaningful changes were made. 1. Very differently: Many substantial changes were made.
Quality of feedback (α = .62)	What percentage of feedback reports were followed by a meeting to discuss the feedback report? Free response What percentage of feedback meetings were conducted with the supervisor present? Free response How long did the typical feedback meeting last? Free response
	During initial feedback meetings, what percentage of the meeting time was characterized by the following behaviors? • Constructive feedback about performance. • Constructive attempts to identify problem causes. • Constructive attempts to develop improvement strategies. • Constructive discussions about future goals. Free response
	After experience with feedback meetings what percent of the meeting time was characterized by the following behaviors? • Constructive feedback about performance. • Constructive attempts to identify problem causes. • Constructive attempts to develop improvement strategies. • Constructive discussions about future goals. Free response
Prior feedback (α = .55)	Frequency of quantitative performance/productivity feedback given to the target unit before ProMES 9. More than once a day 5. Every 2-5 months 1. Never
	Quality of performance/productivity feedback given to the target unit prior to ProMES: Many things go into the quality of the feedback a unit receives. These factors include accuracy, controllability, congruence with overall organizational functioning, timeliness, understandability, and comprehensiveness. Taking all these factors into consideration, how good was the formal and informal feedback the target unit personnel received prior to ProMES? 5. Excellent 3. Adequate 1. Poor
Changes in the feedback system (α = .54)	What percentage of the objectives (products) were substantially changed to obtain formal management approval? (A slight wording change, combining two products into one, or dividing a product into two products are not substantial changes. Adding a new product, dropping a product, or significantly altering the meaning of a product is a substantial change.) Free response What percentage of the indicators were substantially changed to obtain formal approval? (Use the same idea for "substantial" as in changes of objectives above.) Free response What percentage of the contingencies were substantially changed to obtain formal approval? (Substantial here means a change that alters the expected level or other effectiveness scores so that the contingency is really different than it was. A change of 3-5 effectiveness score points would not normally be considered a substantial change.) Free response
	What degree of changes needed to be made to the original system over the first 6 months of feedback? Changes include revising contingencies, changing measures, doing feedback reports differently, etc. Minor changes include changing the expected level on a contingency or adding a graphic to the feedback report. Major changes are done in response to a serious problem such as finding out the indicator data are very different from what was thought. 5. Many major changes had to be made 3. A major change had to be made 1. No changes had to be made
	What degree of changes needed to be made to the original system after the first 6 months of feedback? Changes include revising contingencies, changing measures, doing feedback reports differently, etc. Minor changes include changing the expected level on a contingency or adding a graphic to the feedback report. Major changes are done in response to a serious problem such as finding out the indicator data are very different than what was thought. 5. Many major changes had to be made 3. A major change had to be made 1. No changes had to be made 0. System has not yet run for six months
Trust (α = .89)	Degree of trust the target unit has in management. 5. Very much: Members of the target unit felt that management would never take advantage of them. 3. Moderate: Members of the target unit trusted management would be supportive in most situations but felt they would take advantage of them occasionally. 1. Very little: Target unit members felt that management would take advantage of them at every opportunity.

Table 3 (*continued*)

Variable	Item(s)
	Degree of trust management has in the target unit. 5. Very much: Management felt that the target unit would never take advantage of them. 3. Moderate: Management felt that the target unit would be supportive in most situations but felt that they would take advantage of them occasionally. 1. Very little: Management felt that the target unit would take advantage of them at every opportunity.
Number of personnel in target unit	Approximate number of people in the target unit: Free response
Turnover	What was the average percentage of the target unit personnel annual turnover during the project? Free response
Complexity of the unit (α = .57)	Complexity—Technological. Includes technological and task complexity. Given this definition, how technologically complex was this target unit? 5. Highly complex: The target unit was on the complex end of most of the complexity factors listed above. 3. Moderately complex: The target unit was in the middle of most of the complexity factors listed above. 1. Not complex: The target unit was on the simple end of most of the complexity factors listed above.
	Complexity—Structural. Includes degree of interdependence with other units, number of shifts, and physical separation of target unit personnel. Given this definition, how structurally complex was this target unit? 5. Highly complex: The target unit was on the complex end of most of the complexity factors listed above. 3. Moderately complex: The target unit was in the middle of most of the complexity factors listed above. 1. Not complex: The target unit was on the simple end of most of the complexity factors listed above.
	Complexity—Demands. Includes complex, changing, and sometimes conflicting sets of demands from different sources. Presence of complex internal and external constituencies (e.g. unions and monitoring groups). Given this definition, how complex were the demands on this target unit? 5. Highly complex: The target unit was on the complex end of most of the complexity factors listed above. 3. Moderately complex: The target unit was in the middle of most of the complexity factors listed above. 1. Not complex: The target unit was on the simple end of most of the complexity factors listed above.
Interdependence	Dealing with others: The degree to which the job requires the employee to work closely with other people in carrying out the work activities (including dealings with other organization members and with external organizational "clients'). To what extent did the job require individuals within the group to work with each other? 5. Very much: Dealing with other group members was an absolutely essential and crucial part of doing the job. 3. Moderately: Some dealing with other group members was necessary. 1. Very little: Dealing with other group members was not at all necessary in doing the job.
Management support (α = .82)	At the start of the project (i.e., when the design team started meeting), to what extent did management support the project? Management support is composed of verbal support to the project directors and the target unit, support with organizational resources such as paid employee time and space to work, and publicly stated support of the project to others in the organization. 5. High: Management was willing to invest as many resources and support as needed to ensure the success of the project and helped the project whenever help was needed. 3. Moderate: Management was willing to invest some resources and support in the project, and was helpful in some instances and not in others. 1. Low: Management was unwilling to invest any resources and support in the project and was uncooperative with people involved with the project.
	Once the project was under way, to what extent did management continue to support the project? 5. High: Management continued to be willing to invest as many resources and support as needed to ensure the success of the project and helped the project whenever help was needed. 3. Moderate: Management continued to be willing to invest some resources and support in the project and was helpful in some instances and not in others. 1. Low: Management became unwilling to invest any significant resources and support in the project, and was not helpful when needed.
Centralization (α = .52)	Centralization: The degree to which decision making and authority are centralized or delegated. A completely centralized organization is one where all decision-making authority rests in the hands of a single top manager. A completely decentralized organization is one where every employee has a say in making decisions. To what extent was the structure of the target unit centralized? 5. Highly centralized: Virtually all decision-making authority rested with the supervisor of the target group. 3. Neither: Some important decisions were made by the supervisor and some important decisions were made by target unit personnel. 1. Highly decentralized: All target unit personnel had a say in making virtually all important decisions.
	To what extent was the structure of the local organization centralized? 5. Highly centralized: Virtually all decision-making authority rested with upper management. 3. Neither: Some important decisions were made by the upper management and some important decisions were made by personnel at lower levels of the local organization. 1. Highly decentralized: All personnel had a say in making virtually all important decisions.
Stability of the organization's environment	Stability of the local organization's external environment throughout the course of the project. External environment would include external customer demands, competitors, regulations, the nature of the market, etc. 5. Highly stable: The external environment did not change in meaningful ways during the course of the project. 3. Moderately stable: Some important features of the external environment changed, but many were quite stable during the course of the project. 1. Highly unstable: Most important features of the external environment changed during the course of the project.

Note. The full questionnaire can be found at http://promes.cos.ucf.edu/meta-structural.php

Table 4
Calculated Effect Sizes (ds) for Categorical Moderator Variables

Variable	No. of data points (*k*)	*N*[a]	*d*	Weighted *d*	Corrected *SD*[b]	% Variance due to sampling error	*SE*	95% *CI* Lower	95% *CI* Upper
Overall	83	1647	1.16	1.44	1.44	15.57%	0.16	1.13	1.75
Country									
United States	33	654	0.93	1.15	1.68	11.95%	0.29	0.58	1.72
Germany and Switzerland	23	364	1.18	1.18	0.98	34.31%	0.20	0.78	1.58
Netherlands	13	236	1.39	1.50	0.68	44.64%	0.19	1.13	1.87
Sweden	11	317	1.35	2.21	1.67	9.39%	0.50	1.22	3.20
Function of organization									
Manufacturing	35	517	0.78	1.05	0.86	37.09%	0.15	0.77	1.33
Service (service, education, health care, and military)	42	1018	1.46	1.63	1.69	11.12%	0.26	1.12	2.14
Sales	6	114	1.28	1.45	0.60	48.01%	0.24	0.97	1.93
Type of organization									
Private for profit	44	683	1.00	1.18	1.74	41.91%	0.26	0.67	1.69
Not for profit (includes government)	39	966	1.34	1.62	0.82	9.34%	0.13	1.36	1.88
Type of worker									
Academic and managerial	7	53	2.02	1.74	1.42	47.95%	0.54	0.69	2.79
Blue collar	34	547	1.06	1.54	1.46	17.38%	0.25	1.05	2.03
Clerical	20	437	0.33	0.27	0.75	31.71%	0.17	-0.06	0.60
Technical	22	612	1.78	2.15	1.28	14.51%	0.27	1.62	2.68

[a] *N* = Number of baseline plus feedback periods. [b] The corrected standard deviation of *d* is the *SD* after accounting for sampling variance. It was calculated as the square root of the population variance. The population variance was calculated as the observed variance in *d* minus the sampling variance. The sampling variance was computed using formula 7.38 given by Hunter and Schmidt (2004, p. 288).

of the effectiveness scores from each of the indicators; thus, whereas each indicator has a maximum effectiveness score of +100, the sum of all the effectiveness scores across indicators can be greater than +100. On the horizontal axis, the letter B indicates the baseline periods, and the letter F indicates feedback periods. Recall that the time period varied across studies and depended on the work cycle time and how frequently indicator data were available. In most studies, feedback was given each month.

Results indicated that during baseline, the overall effectiveness score averaged around 50, just above minimum acceptable performance. Overall effectiveness dropped considerably around baseline period 6. This apparent effect is misleading and is caused by studies with different numbers of baseline periods entering the database at different times. Specifically, there were four studies that had six periods of baseline data and thus appear only in the last six baseline periods of the figure. These four studies came from the same overall project that was conducted with four groups of Swedish traffic police officers in the same city. Most of the indicators developed by these researchers concerned spending more time patrolling at places and times that would maximally reduce dangerous driving and hence accidents and injuries. During baseline, the police officers did not patrol this way, and this was reflected in their extremely low overall effectiveness scores. The mean overall effectiveness score during baseline for these four studies was −573. This set of extremely low overall effectiveness scores pulled the mean down substantially for these last six baseline periods.

Once feedback was started, overall effectiveness rose substantially. The figure shows 19 periods of baseline data and 25 periods of feedback data. Going beyond this number of data periods results in values that would be too unstable as a result of the small number of studies. Different combinations of fewer time periods result in the same conclusions as those presented here. These results show large increases in overall effectiveness, with the largest gains occurring in the earlier periods.

Assessing the significance of the change from baseline to feedback with these data was problematic because the data across studies were not directly comparable. The overall effectiveness score for one feedback period in one study was the sum of the effectiveness scores that derive from the indicator scores and their corresponding contingencies during that feedback period. These overall effectiveness scores were comparable across time within a given study. However, the overall effectiveness scores were not interpretable *across* studies. If one work unit consisted of 8 indicators and another unit contained 12, the overall effectiveness scores would tend to be higher in the second unit simply because it contained more indicators. Therefore, the best way to test for changes from baseline to feedback was to calculate the effect size for each study and test whether these are significantly greater than zero. The results of these analyses are presented below, following the discussion of the effect sizes.

Comparison Groups

Data from comparison groups were collected during the same time periods as the ProMES intervention for 18 of the ProMES studies. Data were collected from one or more similar units that did not receive the ProMES intervention. These data were in the form of raw output scores and, thus, were not comparable to ProMES overall effectiveness scores. However, it was possible to calculate

the mean performance for the comparison units during baseline and again during feedback. The mean for the baseline period was 164.76 (SD = 220.34), and the mean for the ProMES period was 168.33 (SD = 224.45), $t(16) = 0.66$, $p = .52$. Thus, the comparison groups showed a minimal, nonsignificant change between baseline and feedback periods.

Overall Effect Size

Although the plot of overall effectiveness scores, as shown in Figure 3, is useful, it has limitations. As noted above, the overall effectiveness scores are not entirely comparable across studies. In addition, the number of studies decreased as we included more time periods. All 83 studies had three total periods of baseline and feedback data. As more periods of feedback and baseline were included, some studies begin to drop out of the average overall effectiveness score. For example, a study with only 5 periods of feedback would appear in the mean overall effectiveness score for the first 5 time periods, but not in the mean for period 6 or later. This influences the interpretation of results because studies that were less successful in improving overall effectiveness were probably ended sooner than those that were more successful. This is supported by the correlation of .21, $t(81) = 1.97$, $p = .03$, one-tailed, between number of feedback periods and the effect size.[2]

A better index of change in effectiveness after feedback is the effect size. An effect size was calculated for each of the 83 studies. A frequency distribution of these effect sizes is shown in Figure 4. These effect sizes ranged from −2.53 to +5.37, with an average effect size of 1.16. This indicates that, on average, productivity under ProMES feedback is 1.16 SD higher than productivity during baseline.

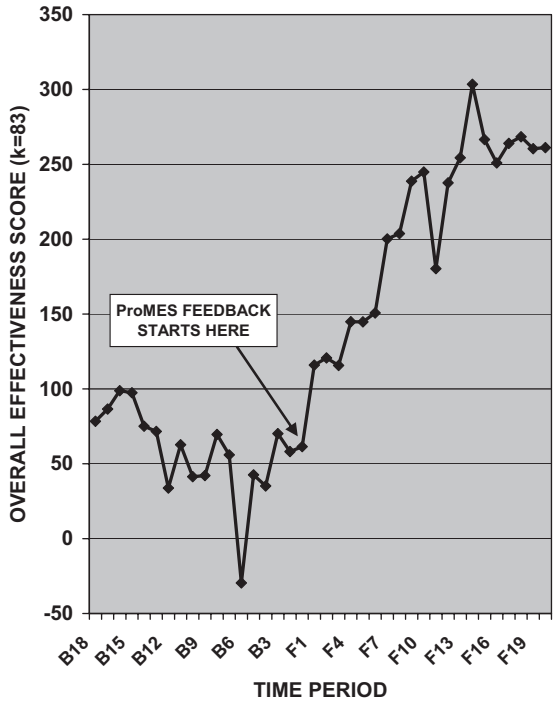

Figure 3. Overall effectiveness scores over time. ProMES = Productivity Measurement and Enhancement System. B represents baseline, and F represents feedback.

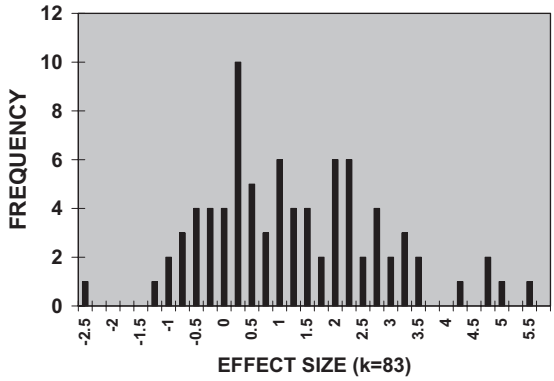

Figure 4. Frequency distribution of effect sizes.

To identify potential outliers, we calculated the sample-adjusted meta-analytic deviancy (*SAMD*) statistic for each project, as described by Huffcutt and Arthur (1995). This procedure identified four outliers. Excluding these four projects, the mean effect size and sample weighted mean effect size were 0.99 and 1.10, respectively. Cortina (2003) recommended that care be taken when eliminating outliers, cautioning that one should only discard outliers when there is overwhelming empirical and substantive justification. In keeping with this, we examined each outlier on a case-by-case basis and found no reason to discard any; therefore, these projects were included in all subsequent analyses.

Hunter and Schmidt (1990) argued, as have others, that studies with larger sample sizes will produce more reliable effect sizes and, because of this, mean effect sizes should also be calculated weighted by sample size. In this research, the analog of sample size would be the number of data periods for each study, the logic being that the more periods of baseline and feedback data, the more reliable is the resulting effect size. The mean effect size weighted by number of periods was 1.44 compared with the unweighted mean of 1.16.[3]

As noted above, the best way to compare the baseline with feedback conditions in these studies was determined to be the examination of effect sizes, because the effect size was interpret-

[2] Given the problems in Figure 3 with the changing number of projects in each feedback period, we also looked at the projects that had at least 10 periods of feedback data. The mean baseline overall effectiveness score for these 45 projects was 98.20. Overall effectiveness means for the first 10 feedback periods were 146.96, 172.30, 179.97, 213.43, 187.35, 182.15, 241.02, 205.90, 229.03, and 245.81. This is essentially the same pattern of means as in Figure 3 for these first 10 feedback periods.

[3] There was one set of five studies (Pritchard et al., 1989) where goal setting and incentives in the form of time off work were added to the ProMES feedback condition. Because goal setting and incentives are not part of ProMES but were added to it, one could argue that these conditions should be removed in calculating mean effect sizes. Mean effect size without these conditions, that is, using just the feedback condition, was 1.07 (compared with 1.16 when they were included) and the weighted mean effect size was 1.25 (compared with 1.44 when included). Thus, the conclusion that the effect sizes were large is the same regardless of whether these conditions are included.

able across studies even when the overall effectiveness score was not. The issue was whether the confidence interval of *d* covered zero. These results are shown in Table 4. As Arthur, Bennett, and Huffcutt (2001) and Whitener (1990) suggested, these confidence intervals are based on the weighted effect sizes. The first row of the table shows the 95% confidence interval for the overall *d.* The interval ranged from a low of 1.13 to a high of 1.75, indicating that the effect sizes were reliably greater than zero. The 95% confidence interval based on the unweighted effect sizes ranged from 0.85 to 1.47.

Another way of understanding the magnitude of our effect sizes is to think of the distribution of performance over time. If we had data in the form of a single index, such as the overall effectiveness score, over a long period of time prior to ProMES, the distribution of these scores would probably be near normal. The mean effect size of 1.16 indicates that the mean overall effectiveness score under ProMES feedback is 1.16 standard deviations higher than the overall effectiveness score under baseline. Being 1.16 standard deviations above the mean of a normal distribution means that 88% of the area under the curve is below this value. Therefore, the average monthly performance under feedback was equivalent to the 88th percentile under baseline. The weighted effect size of 1.44 was the 93rd percentile.

Effects Over Time

The second research question was whether the effects on productivity endure over time. Figure 3 suggests that improvements do last over time. However, because the number of studies decreased as the number of time periods increased, additional analyses were done to address this issue. To examine long-term effects, we selected studies with more than 10 feedback periods. For these studies ($k = 38$), the mean overall effectiveness score for the first 10 periods of feedback and the mean for periods 11 or more were calculated. The mean number of periods greater than 10 was 14.34. The mean overall effectiveness score for the earlier periods was 205.08 ($SD = 240.99$), and the mean for the later periods was 224.74 ($SD = 276.87$). This difference was not significant, $t(37) = 0.93$, $p = .36$. These results indicate that overall effectiveness did not decrease for the later periods.

One potential problem with these results is that in many studies, the overall effectiveness score started fairly low and then increased rapidly, similar to the shape shown in Figure 3. This indicates that the early periods of feedback could have shown lower means because of this lower starting point. The potential problem occurred if overall effectiveness increased early and then decreased later. The relatively low initial scores could mask the decrease in later scores. To deal with this possibility, we again selected the 38 studies with more than 10 feedback periods and split in half the feedback periods from period 11 and above. For example, if a study had 20 feedback periods, periods 11 to 15 were considered as earlier periods, and periods 16 through 20 were considered as later periods. The mean number of periods in the later group was 7.0. The mean overall effectiveness score for the earlier periods was 227.97 ($SD = 261.13$), and the mean for later periods was 222.87 ($SD = 309.97$), $t(37) = -.23$, $p = .82$. Thus, there was a negligible, nonsignificant decrease from these earlier to later periods. These results indicate that the gains in productivity were maintained over time.

Moderators

A critical issue for the present meta-analysis is under what circumstances is the intervention more or less successful. Inspection of Tables 1, 3, and 4 shows substantial variability in the effect sizes, suggesting that moderators are indeed present. Hunter and Schmidt (2004), among others, have recommended testing for the presence of moderators by examining the amount of variance that remains after accounting for artifacts by using the 75% rule. In other words, if 75% or more of the existing variance can be attributed to artifacts, it is unlikely that moderators are present. In the present study, we examined the artifact of sampling error variance and found it to account for only 15.57% of the total variance. As such, an additional 84.43% of the variance remained. This finding provided support for the hypothesis that moderators were present. In fact, the search for moderators of the effectiveness of a ProMES intervention was one of the main reasons for collecting these data over the past 20+ years.

Categorical variables. The third research question was how well the intervention works in different settings. The different settings are operationalized by the categorical moderators shown in Table 4. Results by country show large effect sizes for all four country categories, with Sweden and the Netherlands exhibiting the largest weighted effect sizes. None of the 95% confidence intervals based on the weighted effect sizes for the United States and Germany/Switzerland cover zero. Results for function of the organization show weighted effect sizes to be large for all functions, with service organizations showing the largest effect sizes and manufacturing organizations showing smaller effect sizes. None of the 95% confidence intervals on the weighted effect sizes cover zero. Both private for-profit and not-for-profit organizations showed large weighted effect sizes, with not-for-profit organizations showing the larger effect sizes; neither of the confidence intervals cover zero. Weighted effect sizes for type of worker showed considerable variability. Effects were substantially smaller for clerical workers, by far the lowest weighted mean effect size ($d = .27$) of any subgroup in these categorical variables. The other groups of workers all showed large effect sizes. Only the confidence interval based on weighted effect sizes for the clerical sample covered zero (lower limit = −.06). These results suggest that, with the exception of clerical workers, ProMES shows substantial mean effect sizes in all of these different settings, but there is still considerable variability within subgroups. The confidence intervals were also calculated on the unweighted effect sizes. The conclusions were identical in that none of the confidence intervals covered zero except for clerical jobs, where the lower limit was .00.

Continuous variables. The final research question concerned what factors moderate the effectiveness of the intervention. Such factors are operationalized by the continuous variables found in Table 3. This table shows the internal consistency reliability (alpha) for those variables that were formed from multiple items. Recall that one of the strategies for dealing with the large number of potential moderators was to use composites wherever possible. Alphas for the seven variables that are composites ranged from .52 to .89, with a mean alpha of .64. Some of these alphas are lower than ideal but do indicate that the component variables in the composites are related to each other, thus justifying their inclusion in the composite. The dilemma here was whether to divide the

composites into more homogeneous, internally consistent clusters of fewer items and thereby increase the number of variables and the chances for Type I error or to do the analyses with these composites. We made the decision to proceed with the smaller number of variables. This seems justified because four of the five composites with low (<.70) internal consistency reliabilities were significantly related to effect size (Table 5). This means that although random error may be higher because of the low internal consistency reliability, there is enough systematic variance in the composite to be significantly related to effect size.

To examine the relationships between the continuous moderators and the effect sizes, we conducted two types of analyses: bivariate correlations and weighted least squares (WLS) multiple regression. Table 5 shows the means, standard deviations, and intercorrelations of the continuous moderators and as well as the effect sizes.

The WLS regression was conducted with the inverse of the sampling variance as the weight, as recommended by Steel and Kammeyer-Mueller (2002). These authors argued that WLS is a superior method of detecting continuous moderator variables and has the advantage of identifying the amount of unique variance attributable to each moderator. This is particularly relevant when the moderators are correlated with each other, as they are in this study (see correlations in Table 5). One disadvantage of WLS is that values must be present for all of the moderators for a study to be included in the analysis, which can reduce the sample size. The sample size for the WLS analysis was 47. The sample size for the zero-order correlations ranged from 60 to 82, with a mean of 70.

Table 6 shows the results of the WLS regression. For ease of comparison, the last column of this table repeats the zero-order correlations from Table 5. Overall, the results from the two analyses were quite similar. In 11 of the 12 variables, a variable was either significant or nonsignificant ($p < 05$) in both analyses. The one exception is the amount of prior feedback. It is not significant in the WLS analysis but the correlation ($r = -.42$) is significant.

In addition to the bivariate correlation coefficients between each moderator and the effect size, we also estimated the true relationships by correcting the correlation coefficients for unreliability of the two variables. The results of this analysis are reported in Table 6. The reliability of the moderators was estimated with coefficient alpha when available (see Table 5). Coefficient alpha could not be calculated for five of the moderators because they were assessed using a single item. In these cases, reliability could not be estimated, so no correction was made.

In addition to correcting on the basis of the coefficient alpha, we also computed Rel(d), as suggested by Hunter and Schmidt (2004, p. 295). This is a ratio of estimated true variance of the effect size across studies (excluding sampling error) to the total observed variance, Rel(d) = .84. The correlation coefficients were corrected on the basis of this estimate and the coefficient alpha when available. It should be noted that Rel(d) is not a traditional estimate of reliability in the sense that it does not assess measurement error. As mentioned above, the Rel(d) indicates the proportion of variance across studies that was not attributable to sampling error. The effect of sampling error on the correlations between the effect size and the moderators is similar to that of measurement error; that is, it attenuates the correlations. Thus, Rel(d) is functionally similar to traditional reliability estimates and can be used to correct for the downward biases in the observed correlations between the effect

Table 5
Number of Data Points (ks), Means, Standard Deviations, and Intercorrelations of Study Variables

Correlation	k	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. d	83	1.16	1.55	—												
2. Degree of match	80	4.48	0.75	.44**	—											
3. Quality of feedback	67	0.04	0.60	.43**	.52**	(.62)										
4. Prior feedback	64	−0.01	0.82	−.42**	−.48**	−.17	(.55)									
5. Changes in the feedback system	67	0.07	0.62	−.30*	−.48**	−.27*	.34**	(.54)								
6. Trust	65	3.15	1.09	−.11	.22	.13	−.07	−.29*	(.89)							
7. Number of personnel in target unit	72	22.69	38.04	−.17	−.38**	−.45**	.11	−.06	−.18	—						
8. Turnover	60	8.30	9.77	.02	−.09	−.16	−.01	.03	.40**	.11	—					
9. Complexity of the work unit	65	2.73	0.77	−.05	−.08	−.09	.05	−.16	.12	.37**	.06	(.57)				
10. Interdependence	72	3.49	1.14	−.30**	−.36**	−.42**	.17	−.02	.12	.18	−.01	.25*	—			
11. Management support	71	4.33	0.77	−.09	.01	-.04	.11	−.08	−.04	−.04	.14	−.22	.18	(.82)		
12. Centralization	82	3.26	0.70	.26*	−.05	−.03	−.08	.24	.33**	−.06	.36**	−.18	−.20	.05	(.52)	
13. Stability of the organization's environment	73	3.85	1.27	.04	.03	−.08	.13	−.26*	.10	−.11	.01	−.01	.36**	.52**	−.28*	—

Note. Coefficient alpha is reported for all composite variables along the diagonal.
* $p < .05$. ** $p < .01$.

Table 6
Weighted Least Squares Results

Variable	Unstandardized coefficients		Standardized coefficients			Zero-order r[b]
	B	*SE B*	β	*t*	Corrected r[a]	r w d
Degree of match	0.80	0.30	0.34	2.51*	.48	0.44**
Quality of feedback	0.89	0.37	0.32	2.38*	.59	0.43**
Prior feedback	−0.26	0.29	−0.13	−0.88	−.62	−0.42**
Changes in feedback system	−0.92	0.38	−0.32	−2.39*	−.45	−0.30*
Trust	−0.35	0.18	−0.25	−1.96	−.13	−0.11
Number of personnel in target unit	0.02	0.02	0.16	1.24	−.19	−0.17
Turnover	−0.02	0.02	−0.20	−1.54	.02	0.02
Complexity	0.05	0.26	0.02	0.14	−.07	−0.05
Interdependence	0.42	0.19	0.31	2.23**	−.33	−0.30*
Management support	0.05	0.30	0.03	0.17	−.11	−0.09
Centralization	1.20	0.30	0.48	4.00**	.39	0.26*
Stability of organization's environment	−0.13	0.21	−0.11	−0.61	.04	0.04

Note. The degrees of freedom for the *t* tests of the regression coefficients was 34.
[a] Corrected *r* is the correlation coefficient for *d* and the moderator after correcting for unreliability in each. Reliability of *d* was estimated by the ratio of variance, excluding sampling error, to total variance as recommended by Hunter and Schmidt (2004, p. 295), Rel(*d*) = .84. Reliability of the moderators was estimated by the coefficient alpha for composite variables (see Table 5). When no estimate was available, we assumed a perfect reliability. [b] Zero-order correlations copied from Table 5; r w d represents the zero-order correlation between the variable and the effect size (*d*).
* $p < .05$. ** $p < .01$.

size and the moderators. There are a number of other potential sources that may result in the observed variation in the effect size across studies beyond sampling error; thus Rel(*d*) is likely an overestimate of the reliability of the effect size. Given this, the corrected correlation coefficients are conservative estimates of the true relationships between the variables in the population. These corrected values are shown in the next to last column of Table 6. The corrected correlations are similar to the uncorrected and result in no changes in the conclusions.

Interpretation of the continuous moderators in Tables 4 and 5 is fairly straightforward. One moderator that needs further explanation is degree of match. Degree of match was measured by one item in the questionnaire assessing how closely the development and implementation of the system in a given study matched the process outlined in the first ProMES book (Pritchard, 1990). The item relied on a 5-point Likert scale, and the data showed that all studies received ratings ranging from 3 to 5. A response of 4 indicated that the original process was followed with only minor changes; a response of 3 indicated that a few meaningful changes were made. An example of a study with a score of 4 involved an investigation where management devised the contingencies rather than having the design team create them. An example of a score of 3 was represented by a study in which feedback meetings were not held at a separate location conducive for discussion but on the assembly line while production continued.

Degree of match was strongly related to effect size ($r = .44$), and this finding was confirmed in the WLS analyses. To further explore this variable, mean effect sizes were calculated for studies with ratings of 3, 4, and 5. These results are shown in Figure 5. Of the 80 studies with degree of match data, most (60%) followed the original process, 18.82% had minor changes, and 21.11% had meaningful changes. Figure 5 clearly illustrates that there were major differences in effect size as a function of how well the study followed the original formulation. Studies that followed the original formulation produced a mean effect size, weighted by number of periods, of 1.81; studies with some differences had a mean of 0.60; and studies with substantial differences had a mean of 0.48.

Discussion

Research Question 1

The first research question was whether the ProMES intervention improves productivity. It is clear that overall effectiveness scores improved after the start of ProMES feedback (Figure 3), whereas productivity for the comparison groups showed no change. Effect sizes were large, and the unweighted mean effect size (1.16) was analogous to changing performance from the 50th percentile to the 88th percentile, and for the weighted mean effect size, to the 93rd percentile (1.44).

Research Question 2

The second research question was whether ProMES improvements last over time. The overall results in Figure 3, and the supplemental analyses looking only at studies that extended for more than 10 periods, clearly suggest that they do. These findings are particularly important because it is possible that the novelty of an intervention might temporarily improve productivity but eventually wear off. These effects over time are a significant feature of the ProMES research program. Feedback data were collected over an average of more than 14 periods, over a year for the typical study. In contrast, the 470 effect sizes in the Kluger and DeNisi (1996) feedback meta-analysis had an average feedback period of less than 1 hr.

Research Question 3

The third research question was whether the intervention works in different settings. Results of the categorical analyses indicate

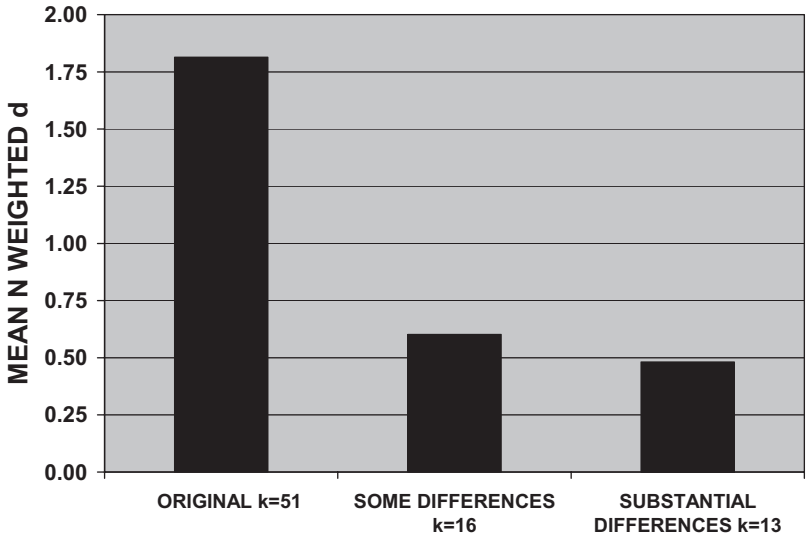

Figure 5. Effect size by degree of match.

that it does. Large mean effect sizes occurred within different countries, organization types, job types, and organizations with different functions. In fact, with one exception, the weighted mean effect size for all of these subgroups was over 1.00. The one exception was for clerical workers, with a weighted mean effect size of 0.27. As it seems to be a reliable finding, it is possible that the type of feedback provided by ProMES is simply not that effective for highly routine clerical jobs. Other researchers have also shown smaller improvements for clerical and other simple tasks (Guzzo et al., 1985; Van Merriënboer, Kester, & Paas, 2006). It may be that because clerical jobs are fairly routine and repetitive, improved task strategies are not really possible, or due to typically short job cycle times, they need more frequent feedback than jobs with longer cycle times. Data on feedback interval were available for 19 of the 20 studies. All but one used a 1-month feedback interval, and one used a 12-week interval. Thus, we were not able to assess empirically whether shorter intervals produced larger effects.

Another interesting finding is that whereas all but one of the mean effect sizes were large, there was considerable variability within each subgroup. For example, the weighted mean effect sizes ranged from 1.05 to 1.63 in the organizational function subgroups. One likely possibility is the presence of moderators within the subgroups. The finding of a number of significant moderators among the continuous variables supports this interpretation and points to one area of future research. In addition, other studies have identified moderators of some of the categorical (Harrell & Pritchard, 2007) and continuous variables (Watrous et al., 2006).

Although ProMES seems to work in many settings, there are some situations where ProMES is not appropriate. One such situation is that in which the nature of the work changes so often that doing and redoing the measurement system is not cost-effective. This most often occurs when objectives and indicators change frequently. This could happen if the nature of the technology, customer requirements, or the ways in which the work unit is structured change regularly. It is often less of an issue if contingencies change frequently. For example, where the work involves doing a series of well-defined projects over time, such as consulting, construction, and software design, the specific project may change frequently but the basic objectives and indicators do not change. Quality of work, timeliness, meeting customer requirements, and working efficiently stay the same; however, the measures often remain the same. What may change are the contingencies because relative importance and what is considered as minimally acceptable performance may vary for different projects. Changing the contingencies in such situations can be done quickly, typically in a meeting of an hour or so. In other words, if contingency changes are needed frequently, this can often be accommodated.

Research Question 4

The fourth research question dealt with the factors influencing the effectiveness of the intervention. The first variable measured how closely the study followed the original ProMES formulation. This degree of match variable was highly related to effect size. It is clear that when the original process is followed, the effect sizes are very large, and they drop off dramatically when the intervention deviates from this original process. Another interpretation of this result is that the situational factors hindering the full ProMES process could be the cause of the intervention being less successful. For example, a management team that is not comfortable with unit members developing contingencies might allow for less autonomy in many situations. This could decrease the effectiveness of a participative intervention like ProMES.

The next set of moderators dealt with feedback. Quality of feedback was strongly related to effect size ($r = .43$). The quality of feedback composite measured what happened during the feedback meetings. The findings show that how those meetings are conducted is highly related to the effectiveness of the intervention.

Level of changes in the feedback system, was, as predicted, negatively related to productivity improvement ($r = -.30$). This composite combines the changes made by management during the approval and the changes made to the system after it was imple-

mented. The need for such changes by management suggests that the design team did not fully understand their task, and the need for such changes after implementation suggests that they were not initially able to come up with an optimal measurement system. This suggests that design teams that can devise an optimal system from the start will benefit more from it compared with teams that cannot.

Amount of prior feedback was also negatively related to *d* in the correlation analyses ($r = -.42$), but was not significant in the WLS analyses. Presumably, variance accounted for by this variable was shared with other moderators, or the relationship is simply different in the reduced sample of studies in the WLS analysis ($k = .47$) compared with the sample for the bivariate correlation results ($k = .64$).

Most of the unit-level moderators showed no significant relationship with the effect size. The one exception was interdependence ($r = -.30$), which supports the idea that feedback focused on outputs, such as typical ProMES indicators, rather than processes may be less effective for highly interdependent units (Smith-Jentsch, Zeisig, Action, & McPherson, 1998). We argued that the feedback meetings should provide this process feedback, and Garafano et al. (2005) found that when feedback meetings were conducted well, this reduced the negative relationship between interdependence and effect size.

It is also instructive to look at some of the variables that were not related to the effect size. Although this gets close to the problematic issue of accepting the null hypothesis, we can still say in a descriptive sense that some variables do not seem to be related to the effectiveness of the intervention. For example, we hypothesized that trust, number of people in the unit, turnover, complexity, management support, and stability of the organization's environment would be related to productivity improvement, but none of these variables were. This suggests that variation in these variables is not critical to the success of a study. One variable that especially surprised us was management support. A possible explanation for this is that management support may have had its effects before a study ever started. If management was not supportive, the project would either not be started at all or would be stopped before feedback starts.

It is also interesting that turnover is not related to effect size. We suspect that this has to do with the nature of the intervention. It is possible that the level of information shared during the development of the system and the resulting feedback meetings made it easier to train people new to the unit. Another possibility is that greater commitment to high performance in the unit had positive effects on the performance of new people. One or both of these processes could overcome the negative effects of having to train and socialize new personnel as a result of turnover.

Implications for Theory

Reasons for large effects. A major question for theory is why ProMES has such large effects on productivity. One possible explanation comes from NPI and the Pritchard–Ashwood theories. ProMES was designed to simultaneously influence as many of the variables effecting the motivational process as possible. In contrast, most other interventions influence only a subset of these motivational variables. This part–whole difference may be why ProMES produces such large improvements. Figures 6 and 7 make this point graphically. The center column of Figure 6 shows the components of the theory. The effects of developing the ProMES measurement system are shown on the left side of the figure, and the effects of implementing feedback are shown on the right side. Starting first with system development, the ProMES indicators are the operationalization of the Results; objectives are the mechanism used to help identify the Results. This is indicated by the dotted arrows from objectives and from indicators to Results. When indicators are developed, this clarifies which results are measured by the organization and how this measurement is done. ProMES contingencies are the operationalization of result-to-evaluation connections and relate levels of each indicator (result) to the level of overall effectiveness (the evaluation).

Once ProMES is developed, feedback begins. As is shown on the right side of the figure, the feedback itself gives concrete information on how much of each result was done for the period. This is the score on the indicator and is shown by the arrow to the -results box. The feedback report also indicates how positive that amount of the result was: the effectiveness score for the indicator. The effectiveness score is an evaluation. Because unit members have information on both the results (indicators) and the evaluations (effectiveness scores), this also helps to clarify result-to-evaluation connections.

Having ProMES feedback meetings over time has effects on several motivation components. Feedback meetings are used to plan improvement strategies and evaluate strategies that have been tried in the past. Part of this discussion involves making plans to overcome situational constraints on performance (Finkelstein & Hambrick, 1990; Peters & O'Connor, 1980) and improve task strategies (e.g., Earley, Connolly, & Ekegren, 1989). This discussion and improvement of task strategies in the feedback meetings also changes actions-to-results connections. As new strategies are tried and improved, actions and actions-to-results connections continue to change.

Feedback over time has some positive effect on outcomes. If the evaluation system is perceived as fair, performing well should lead to more positive outcomes. Such outcomes may lead to feelings of accomplishment, which may not occur if the evaluation system was considered unfair (Cropanzano & Greenberg, 1997; Folger, Konovsky, & Cropanzano, 1992). If performance improves, other outcomes such as recognition and salary could also increase. ProMES can also change the evaluations-to-outcomes connection. If by implementing ProMES, the evaluation system is clearer, then the connections between evaluations and outcomes become clearer.

It is also possible that a given outcome could satisfy more needs under ProMES feedback than before. For example, if the evaluation system is perceived as accurate, a pay raise based on that evaluation system could lead to satisfaction of needs for achievement in addition to just needs for money. This is an example of changing the outcomes-to-need satisfaction connection. Thus, between (a) development of the measurement system, (b) receiving feedback, and (c) using the feedback to make improvements, there are direct links between ProMES components and the entire motivational chain.

In contrast, Figure 7 illustrates how more traditional interventions influence the motivational components. For example, training is geared to teaching the person how to use their actions to produce results that are desired by the organization. Thus, training

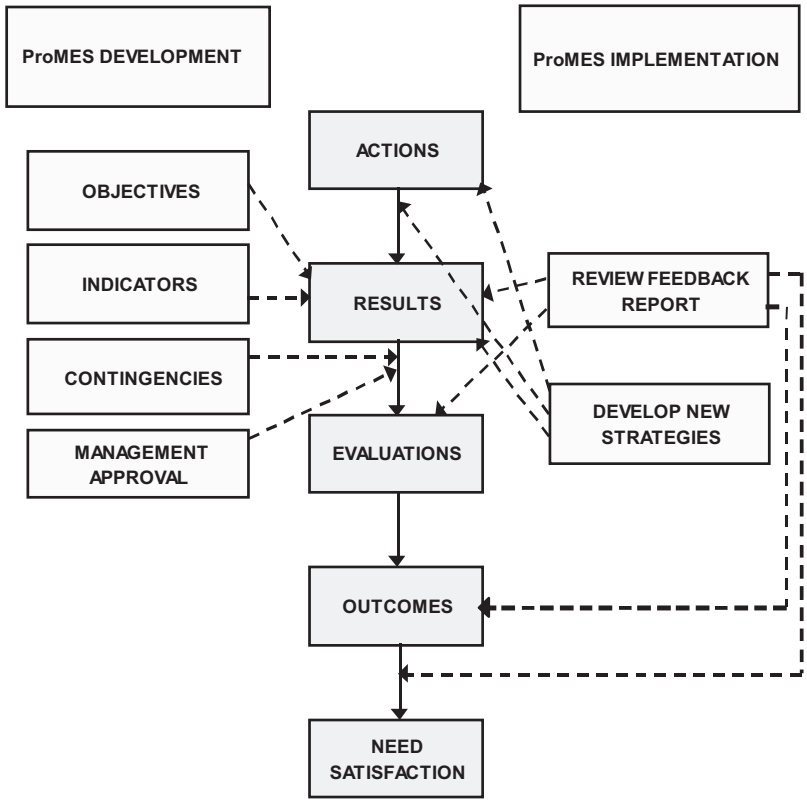

Figure 6. Effects of Productivity Measurement and Enhancement System (ProMES) intervention on motivation components.

effects actions-to-results connections and helps clarify the expected results. Formal goal setting changes results-to-evaluations connections by defining the point (the goal) where the evaluations increases dramatically (Naylor et al., 1980; Naylor & Ilgen, 1984). When not meeting the goal, the evaluation is not particularly positive; at or above the goal, the evaluation is high. Feedback on output levels influences results and helps clarify the actions-to-results connection. Performance appraisal in the form of ratings by others clarifies how outputs and other behaviors lead to evaluations and thus clarifies results-to-evaluations connections and evaluations. Finally, incentives that provide rewards for superior performance influence the evaluations-to-outcomes connection and outcomes.

Thus, each of these interventions affects only a subset of the motivational process. If the other motivational variables are sufficiently high, increasing the one or two motivational components these interventions are geared toward will have a positive effect on motivation and performance. However, if any of the other variables are low, improving one or two motivational components in isolation will have much less effect. Training will have little impact if the people do not have a good feedback system on the job that tells them what level of outputs (results) they are generating. In contrast, ProMES can have some effect on all the components and thus have a greater potential effect on motivation and performance.

Support for the underlying theory. One cannot say that the ProMES research program tests the NPI and Pritchard–Ashwood theories, but the results are at least consistent with the theories. The theories suggest that if all of the connections between actions, results, evaluations, outcomes, and need satisfaction are strong, motivation will be high. ProMES was designed to accomplish exactly that. Therefore, the fact that the intervention produces large effects on productivity is consistent with the theory. The other finding that is consistent with the theory comes from the degree of match data. The theory suggests that if any of the connections are low, motivation will be low. The fact that omitting or changing parts of the methodology seems to result in much lower effect sizes is consistent with this aspect of the theory as well.

Combining theory with practice: Evidence-based management. Another implication for both theory and practice is that it is possible to combine theory with practice. This was a major goal of the ProMES research program and it seems to have been successful. The major advantage of a theory-based intervention is that it guides decision making about the intervention. For example, in deciding how to develop contingencies, the theory made it clear that this should be done in a way that the effectiveness scores must be interpretable across indicators. For example, an effectiveness score of +20 must be equally valuable to the organization regardless of whether the indicator was being considered. This had a major effect on the design of the steps used in developing the contingencies.

The ProMES research program can be seen as an example of the concept of evidence-based management (Pfeffer & Sutton, 2006;

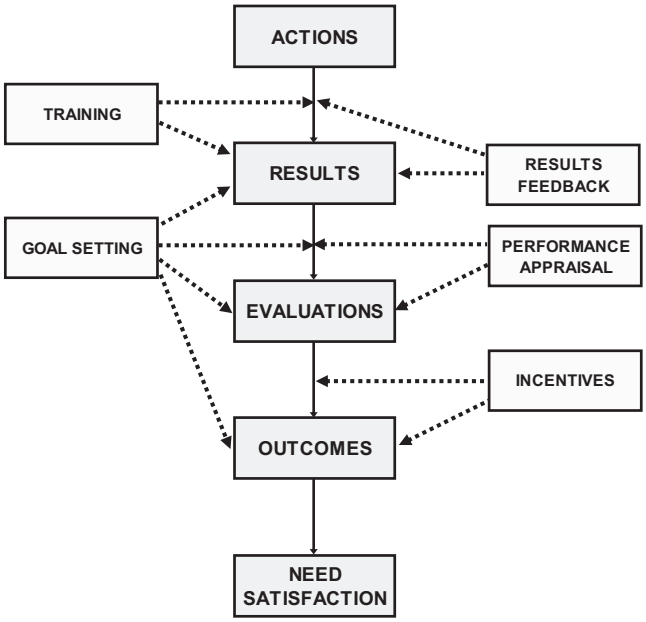

Figure 7. Effects of other interventions on motivation components.

Rousseau, 2005, 2006; Rousseau & McCarthy, 2007; Walshe & Rundall, 2001). The basic idea is that managers should use techniques and interventions based on evidence of their success rather than on other factors, such as what other organizations do. Rousseau and McCarthy (2007) argued that evidence-based management offers the promise of better outcomes from management decisions; continued learning by managers; and closer ties between scholars, educators, and practitioners. Because ProMES is based on sound theory, and there is clear evidence of its effectiveness in organizations, ProMES is an example of an intervention that is evidence-based.

Importance of task strategies. A final implication for theory comes from experience working with ProMES. On the basis of user feedback, the large effects of ProMES seem to be achieved without the expenditure of much additional effort (Minelli, 2005). What changes is how that effort is expended. This suggests that task strategies may be a much more important area to explore than has been done in the past. There has been some empirical work on strategy (e.g., Earley et al., 1989; Kanfer & Ackerman, 1989; Schmidt, Kleinbeck, & Brockmann, 1984; Wood & Locke, 1990) as well as some conceptual work, especially in the area of action theory (e.g., Frese & Sabini, 1985; Hacker, Volpert, & von Cranach, 1982; Heckhausen, 1991; Kuhl & Beckmann, 1985; Semmer & Frese, 1985). However, our anecdotal findings suggest that task strategies are a very important issue for future theory and research.

Implications for Practice

Without a doubt, the most important implication for practice is that we can have a major effect on productivity. Being able to improve productivity so that the mean is now equal to what was the 88th (unweighted) or 93rd (weighted) percentile during baseline is a major effect. Although we have much less data on this, what data are available suggest this can be accomplished with an increase, or at least no decrease, in job satisfaction and with a decrease in stress (Pritchard, Paquin, et al., 2002).

We can also consider what the data tell us about how ProMES should be implemented to maximize its effects on productivity. However, we need to interpret these results with caution because we cannot assign causality on the basis of the results of this meta-analysis. With that caveat in mind, the data suggest that the intervention should be carried out as described in the original formulation (Pritchard, 1990); feedback should be of high quality, as defined by the items in the quality of feedback measure; and the design team and management should develop the system as completely and accurately as possible to avoid changes once it has been implemented. As for situational factors, the intervention seems to be most successful where the interdependence in the unit is low, and the organization is more centralized. Type of organization is not critical to its success, nor is type of job, with the exception of clerical jobs. ProMES was successful in all countries where it has been used. However, all of these countries are Western cultures with successful economies. How it might work in other types of economies and cultures is not yet known.

As a final point in this section, it is instructive to look at how much overall variance in d was accounted for by the moderators. The multiple R for the WLS analysis was .77. However, the number of predictors was large for the number of studies. A better estimate is provided by the estimated multiple R after shrinkage. To calculate this, we used the formula suggested by Cohen, Cohen, West, and Aiken (2003, p. 84): $\tilde{R}^2 = 1 - (1 - R^2)[(n - 1)/(n - k - 1)]$. This value was .45. Thus, with the adjusted R, the moderators accounted for 45% of the variance in d.

Limitations

There are a number of limitations of this research. One is that we do not have available all of the ProMES studies that have been done. Although attempts were made over the past 18 years to obtain data from these studies, this was not entirely successful. However, there is no reason to expect that the studies not included were systematically different from those that were included. Compared with most meta-analyses of organizational interventions, it is likely that this review includes a higher proportion of the possible studies and a much higher proportion of the unpublished studies.

There are other limitations as well. Selection bias is possible because work units using ProMES were not randomly selected. In all cases, personnel in management decided which units would be used. In many cases, the supervisors and occasionally the people in the units had to be willing to try the intervention. This may have resulted in greater improvements than would have been the case for units where supervision or members were opposed to the intervention. Another source of potential error is that the intervention is not always implemented in the same way. Constraints in the setting result in differences in implementation, different facilitators use different styles, and different design teams and the units they represent make different decisions about how to design the feedback reports. However, both the selection bias and lack of an invariant intervention are characteristics of most interventions used in organizations, and there is no reason to expect they would cause more problems with ProMES than with other interventions.

Another issue to consider is a possible Hawthorne effect. Because ProMES represents a substantial increase in the amount of

attention the work unit receives, this attention alone may produce increases in productivity. However, this seems unlikely because in the ProMES study design, the increased attention occurs when the design team and the rest of the unit are developing the system. The first baseline data are not available until after the measures are finalized. This means that by the time collection of baseline data starts, the unit has been working under the increased attention for some time, usually for months. Therefore, any Hawthorne effect should occur prior to the start of the baseline period and thus not influence the effect size.

A series of limitations with the moderator analyses were also present. The first concerned the meta-analysis questionnaire itself. The major advantage of having investigators complete the questionnaire is that they are most familiar with the study and can provide much richer data than a typical meta-analysis coder. However, the disadvantage is not being able to train the raters and assess interjudge reliability. In addition, many of those completing the questionnaire were not native English speakers. One effect of this potential unreliability is to add random error to the questionnaire data. Such random error would be due to investigators interpreting some of the items differently or using different response styles. If such errors were random, the effect would be to increase error variance and decrease the size of the relationships of the moderators with the effect size. Thus, any such random error effects would make the obtained results conservative.

The other possibility is systematic error. If an investigator believes that there is a relationship between productivity gain (*d*) and a given moderator, this could influence their responses. They could report values of the moderator on the basis of the actual productivity gain and his or her expectations, even if unintentionally. Although such systematic bias is possible, for it to have influenced the results, most or all of the investigators would have to have had the same bias, which seems unlikely.

Three other potential limitations could have increased random error. One is that the questionnaire was sometimes completed months after the study began and, in some cases, even after the study was over. Another problem is that some of the questions may have been difficult for the researcher to assess. For example, the researcher may not have a clear picture of how much trust there was between management and the unit personnel. If the respondent provided responses well after the fact or when unsure of the correct response, the effect would likely be random error, which would have made the observed results more conservative and the likelihood of Type II error greater. The same effects would be expected from the relatively low internal consistency reliabilities of some of the composite moderator variables.

Another concern is the use of single-item measures for several of the moderators. A common criticism of single-item measures is that they are unreliable (Wanous, Reichers, & Hudy, 1997). As in the case of the other sources of random error, this unreliability would reduce the size of any relationships, making obtained results more conservative and increasing the chances of Type II error. In support of single-item measures, prior studies have found that single-item measures can be psychometrically comparable to multiple-item measures (Gardner, Cummings, Dunham, & Pierce, 1998; Wanous et al., 1997).

Another potential problem is range restriction in the moderators. Because management had to approve the initiation of a study, studies were conducted only in organizations where management was supportive of such an intervention. For example, organizations where trust between management and unit personnel was low might not be willing to try an intervention such as ProMES. This could have produced range restriction, because organizations where trust was low would not appear in the database. Any effects from such restriction would likely have reduced the observed relationships with the effect size. This suggests that there may be some variables that did influence the effectiveness of the intervention, but our results cannot detect these relationships.

Conclusions

The single most important overall conclusion from this research is that productivity improvements were very large with the ProMES intervention. Such large improvements suggest that there is great potential in organizations that is not being utilized. In one sense, this is a tragedy because untapped potential is wasted. The problem does not lie with the individuals themselves because the individuals in the studies were the same before and after the productivity improvements. The difficulties must be related to the jobs and the organizational practices. People are working in jobs that severely limit what they can contribute. In another sense, this state of affairs is not a tragedy, because it offers a hopeful challenge for the future. These findings suggest that changes can be made that unlock this huge potential. The task is to discover ways to make these changes. ProMES is one approach, but other approaches can be developed to use this potential and improve the work lives of people in organizations.

References

*An asterisk represents articles included in the meta-analysis.

Abbasi, S. M., & Hollman, K. W. (2000). Turnover: The real bottom line. *Public Personnel Management, 29,* 333–342.

*Agrell, A., & Malm, K. (2002). ProMES in a Swedish traffic police department and its effects on team climate. In R. D. Pritchard, H. Holling, F. Lammers, & B. D. Clark (Eds.), *Improving organizational performance with the Productivity Measurement and Enhancement System: An international collaboration* (pp. 53–68). Huntington, New York: Nova Science.

Anderson, N. R., & West, M. A. (1994). *The Team Climate Inventory: Manual and users' guide.* Windsor, United Kingdom: Assessment Services for Employment NFER-Nelson.

Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 8,* 234–245.

Arthur, W., Jr., Bennett, W., Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS.* Mahwah, NJ: Erlbaum.

Balcazar, F. E., Hopkins, B. L., & Suarez, Y. (1986). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7,* 65–89.

Balzer, W. K., Doherty, M. E., & O'Connor, R., Jr. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 410–433.

Beadles, N. A., II, Lowery, C. M., Petty, M. M., & Ezell, H. (2000). An examination of the relationships between turnover functionality, turnover frequency, and organizational performance. *Journal of Business and Psychology, 15,* 331–387.

Bobko, P., & Colella, A. (1994). Employee reactions to performance standards: A review and research propositions. *Personnel Psychology, 47,* 1–29.

*Bonic, I. (1995). ProMES and computer service technicians: An Australian application. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 119–139). New York: Praeger.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica, 45,* 223–241.

Brown, T. C., & Latham, G. P. (2002). The effects of behavioral outcome goals, learning goals, and urging people to do their best on an individual's teamwork behaviour in a group problem-solving task. *Canadian Journal of Behavioural Science, 34,* 276–285.

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* (pp. 63–130). Chicago: Rand McNally College.

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology, 83,* 615–633.

*Clark, K. (1999). *Using ProMES as a strategic planning approach.* Unpublished doctoral dissertation, Texas A&M University.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfillment. *Journal of Occupational Psychology, 53,* 39–52.

Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods, 6,* 415–439.

Cropanzano, R., & Greenberg, J. (1997). Progress in organizational justice: Tunneling through the maze. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 12, pp. 317–372). New York: Wiley.

Dalton, D. R., Todor, W. D., Spendolini, M. J., Fielding, G. J., & Porter, L. W. (1980). Organization structure and performance: A critical review. *Academy of Management Review, 5,* 49–64.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8,* 377–383.

David, J. H. (2005). Identifying the factors that contribute to the effectiveness of the productivity measurement and enhancement system (ProMES). *Dissertation Abstracts International, 65,* 3755.

Dipboye, R. L., & de Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. *Journal of Applied Psychology, 66,* 248–251.

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87,* 611–628.

Earley, P. C., Connolly, T., & Ekegren, G. (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *Journal of Applied Psychology, 74,* 24–33.

Erez, M., & Kanfer, F. H. (1983). The role of goal acceptance in goal setting and task performance. *Academy of Management Review, 8,* 454–463.

Ergeneli, A., Ari, G. S., & Metin, S. (2007). Psychological empowerment and its relationship to trust in immediate managers. *Journal of Business Research, 60,* 41–49.

Finkelstein, S., & Hambrick, D. C. (1990). Top-management-team tenure and organizational outcomes: The moderating role of managerial discretion. *Administrative Science Quarterly, 35,* 484–503.

Fisher, C. D., & Gitelson, R. (1983). A meta-analysis of the correlates of role conflict and ambiguity. *Journal of Applied Psychology, 68,* 320–333.

Folger, R. (1977). Distributive and procedural justice: Combined impact of "voice" and improvement on experienced inequity. *Journal of Personality and Social Psychology, 35,* 108–119.

Folger, R., Konovsky, M. A., & Cropanzano, R. (1992). A due process metaphor for performance appraisal. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational Behaviorism,* (Vol. 14, pp. 129–177). Greenwich, CT: JAI Press.

Frenkel, A., Tam, M., Korczynski, M., & Shire, K. (1998). Beyond bureaucracy? Work organization in call centers. *The International Journal of Human Resource Management, 9,* 957–979.

Frese, M., & Sabini, J. (1985): *Goal directed behaviour: The concept of action in psychology.* London: Erlbaum.

Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H. C. Triandis, M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 4, 2nd ed., pp. 271–340). Palo Alto, CA: Consulting Psychologists Press.

Garafano, C., Kendall, D., & Pritchard, R. D. (2005, April). *Group interdependence, type of feedback, and changes in productivity.* Poster presented at the 20th annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical examination. *Educational and Psychological Measurement, 58,* 898–915.

Guzzo, R. A., Jette, R. D., & Katzell, R. A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. *Personnel Psychology, 38,* 275–291.

Hacker, W., Volpert, W., & von Cranach, M. (Eds.). (1982). *Cognitive and motivational aspects of action.* Amsterdam: North-Holland.

Harrell, M., & Pritchard, R. D. (2007, April). *Trust and productivity improvement: A cross-cultural analysis.* Paper presented at the Society for Industrial and Organizational Psychology Conference, New York.

Heckhausen, H. (1991). *Motivation and action.* Berlin: Springer.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* London: Sage.

Holling, H., Lammers, F., & Pritchard, R. D. (Eds.). (1998). *Effektivität durch partizipatives Produktivitätsmanagement.* [*Effectiveness through participative productivity management*]. Göttingen, Germany: Hogrefe.

House, R. J., Hanges, P. J. Javidan, M. Dorfman, P. W., & Gupta, V. (Eds.). (2004). C*ulture, leadership, and organizations: The GLOBE study of 62 societies.* Thousand Oaks, CA: Sage.

*Howell, C., Jones, S. D., & Hood, R. L. (1995). ProMES in a service setting. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 170–189). New York: Praeger.

Huffcutt, A. I., & Arthur, W., Jr. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80,* 327–334.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks: Sage.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

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.

Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Decision Processes, 36,* 16–78.

*Janssen, P., van Berkel, A., & Stolk, J. (1995). ProMES as part of a new management strategy. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 43–61). New York: Praeger.

*Jones, S. D. (1995a). ProMES in a small manufacturing department:

Results of feedback and user reactions. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 81–93). New York: Praeger.

*Jones, S. D. (1995b). ProMES with assembly line work groups: It is more than just a technology. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 94–116). New York: Praeger.

* Jones, S. D., McCarthy, P. M., Wagner, S. L. & Hein, M. B. (2002). Effectiveness measurement for a knowledge work group of industrial and organizational psychologists. In R. D. Pritchard, H. Holling, F. Lammers, & B. D. Clark (Eds.), *Improving organizational performance with the Productivity Measurement and Enhancement System: An international collaboration.* (pp. 69–88). Huntington, NY: Nova Science.

Kanfer, R. (1990). Motivation theory and industrial/organizational psychology. In M. D. Dunnette & L. Hough (Eds.), *Handbook of industrial and organizational psychology: Vol. 1. Theory in industrial and organizational psychology* (pp. 75–170). Palo Alto, CA: Consulting Psychologists Press.

Kanfer, R. (1992). Work motivation: New directions in theory and research. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 7, pp. 1–53). London: Wiley.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition [Monograph]. *Journal of Applied Psychology, 74,* 657–690.

Kanfer, R., Ackerman, P. L., Murtha, T. C., Dugdale, B., & Nelson, L. (1994). Goal setting, conditions of practice, and task performance: A resource allocation perspective. *Journal of Applied Psychology, 79,* 826–835.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65,* 681–706.

Keil, C. T., & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin, 127,* 673–697.

Kirkman, B. L., & Shapiro, D. L. (2001). The impact of team members' cultural values on productivity, cooperation, and empowerment in self-managing work teams. *Journal of Cross-Cultural Psychology, 32,* 597–617.

Klein, K. J., & Sorra, J. S. (1996). The challenge of innovation implementation. *Academy of Management Review, 21,* 1055–1080.

*Kleingeld, A., Van Tuijl, H., & Algera, J. A. (2004). Participation in the design of performance management systems: A quasi-experimental field study. *Journal of Organizational Behavior, 25,* 831–851.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254–284.

Kopelman, R. E. (1986). Objective feedback. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 119–144). Lexington, MA: Heath.

Korczynski, M. (2001). The contradictions of service work: Call centre as customer-oriented bureaucracy. In A. Sturdy, I. Grugulis, & H. Willmott (Eds.), *Customer service: Empowerment and entrapment* (pp. 79–101). Basingstoke, England: Palgrave.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology, 50,* 569–598.

Kuhl, J., & Beckmann, J. (1985). Historical perspectives in the study of action control. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 89–100). Berlin: Springer-Verlag.

Latane, B. (1981). The psychology of social impact. *American Psychologist, 36,* 343–356.

Latane, B., & Nida, S. (1980). Social impact theory and group influence: A social engineering perspective. In P. Paulus (Ed.), *Psychology of group influence.* Hillsdale, NJ: Erlbaum.

Latane, 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,* 823–832.

Latham, G. P. (2007). *Work motivation: History, theory, research and practice.* Thousand Oaks, CA: Sage

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., & Pinder, C. C. (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology, 56,* 485–516.

Lind, E. A., Kanfer, R., & Earley, P. C. (1990). Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments. *Journal of Personality and Social Psychology, 59,* 952–959.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice.* New York: Plenum.

Lindell, M. K. (1976). Cognitive and outcome feedback in multiple-cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 739–745.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57,* 705–717.

Locke, E. A., & Schweiger, D. M. (1979). Participation in decision-making: One more look. *Research in Organizational Behavior, 1,* 265–340.

London, M. (2003). *Job feedback: Giving, seeking, and using feedback for performance improvement* (2nd ed.). Mahwah, NJ: Erlbaum.

Massie, J. L. (1965). Management theory. In J. G. March (Ed.), *Handbook of organizations* (pp. 387–422). Chicago: Rand McNally.

Mayer, R. C., Davis, J. H., & Schoorman, D. R. (1995). An integrative model of organizational trust. *Academy of Management Review, 20,* 709–734.

McElroy, J. C., Morrow, P. C., & Rude, S. N. (2001). Turnover and organizational performance: A comparative analysis of the effects of voluntary, involuntary, and reduction-in-force turnover. *Journal of Applied Psychology, 86,* 1294–1299.

*Miedema, H., Thierry, H., & van Oostveen, F. (1995). ProMES in a small oil distribution company. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 229–242). New York: Praeger.

*Minelli, M. (2005). *Leistungssteigerung dank aktivitätsorientierter Verkaufsführung: Ergebnisse aus einer ProMES-Fallstudie mit quasi-experimentellem Längsschnitt-Design* [*Performance improvement through activity-based sales management: Results from a ProMES case study with a longitudinal control-group design*]. Doctoral dissertation, University of Bern, Switzerland.

*Minelli, M., Walliser, F., Tschan, F., Herzog, W., & Semmer, N. K. (2002). ProMES in a Swiss school: Effects of priority information in feedback. In R. D. Pritchard, H. Holling, F. Lammers, & B. D. Clark, (Eds.), *Improving organizational performance with the Productivity Measurement and Enhancement System: An international collaboration* (pp. 87–104). Huntington, New York: Nova Science.

Mitchell, T. R., & Daniels, D. (2003). Motivation. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 225–254) Hoboken, NJ: Wiley.

Mobley, W. H. (1982). *Employee turnover, causes, consequences, and control.* Reading, MA: Addison-Wesley.

Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance, 2,* 183–200.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance*

appraisal: Social, organizational, and goal-based perspectives. Thousand Oaks, CA: Sage.

Naylor, J. C., & Ilgen, D. R. (1984). Goal setting: A theoretical analysis of a motivational technology. *Research in Organizational Behavior, 6,* 95–140.

Naylor, J. C., Pritchard, R. D., & Ilgen, D. R. (1980). *A theory of behavior in organizations.* New York: Academic Press.

Paquin, A. R. (1997). A meta-analysis of the Productivity Measurement and Enhancement System. *Dissertation Abstracts International, 58,* 3953.

Parker, C. P., Baltes, B. B., Young, S. A., Huff, J. W., Altmann, R. A., Lacost, H. A., & Roberts, J. E. (2003). Relationships between psychological climate perceptions and work outcomes: A meta-analytic review. *Journal of Organizational Behavior, 24,* 389–416.

Peters, L. H., & O'Connor, E. J. (1980). Situational constraints and work outcomes: The influences of a frequently overlooked construct. *Academy of Management Review, 5,* 391–397.

Pfeffer, J., & Sutton, R. I. (2006). *Hard facts, dangerous half-truths, and total nonsense.* Boston: Harvard Business School Press.

Price, J. L. (1989). The impact of turnover on the organization. *Work and Occupations, 16,* 461–473.

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* (Vol. 3, 2nd ed., pp. 443–471). Palo Alto, CA: Consulting Psychologists Press.

*Pritchard, R. D. (Ed.). (1995). *Productivity measurement and improvement: Organizational case studies.* New York: Praeger.

Pritchard, R. D., & Ashwood, E. L. (in press). *Managing motivation: A manager's guide to diagnosing and improving motivation.* New York: Erlbaum/Psychology Press.

Pritchard, R. D., Holling, H., Lammers, F., & Clark, B. D. (Eds.). (2002). *Improving organizational performance with the Productivity Measurement and Enhancement System: An international collaboration.* Huntington, NY: Nova Science.

*Pritchard, R. D., Jones, S. D., 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 Series, 73,* 337–358.

*Pritchard, R. D., Jones, S. D., 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., Kleinbeck, U. E., & Schmidt, K. H. (1993). *Das Management-system PPM: Durch Mitarbeiterbeteiligung zu höherer Produktivität.* [*The PPM Management System: Employee participation for improved productivity*]. Munich, Germany: Beck.

Pritchard, R. D., Paquin, A. R., DeCuir, A. D., McCormick, M. J., & Bly, P. R. (2002). Measuring and improving organizational productivity: An overview of ProMES, The Productivity Measurement and Enhancement System. In R. D. Pritchard, H. Holling, F. Lammers, & B. D. Clark (Eds.), *Improving organizational performance with the Productivity Measurement and Enhancement System: An international collaboration* (pp. 3–50). Huntington, NY: Nova Science.

*Pritchard, R. D., Watson, M. D., Kelly, K., & Paquin, A. (1998). *Helping teachers teach well: A new system for measuring and improving teaching effectiveness in higher education.* San Francisco: New Lexington Press.

Pritchard, R. D., Youngcourt, S. S., Philo, J. R., McMonagle, D. C., & David, J. H. (2007). Priority information in performance feedback. *Human Performance, 20,* 61–83.

*Przygodda, M., Kleinbeck, U., Schmidt, K., & Beckmann, J. (1995). Productivity measurement and enhancement in advanced manufacturing systems. In R. D. Pritchard (Ed.), *Productivity measurement and improvement: Organizational case studies* (pp. 62–80). New York: Praeger.

Rafferty, J., & Tapsell, J. (2001). Self-managed work teams and manufacturing strategies: Cultural influences in the search for team effectiveness and competitive advantage. *Human Factors and Ergonomics in Manufacturing, 11,* 19–34.

Rodgers, R., & Hunter, J. E. (1991). Impact of management by objectives on organizational productivity. *Journal of Applied Psychology, 76,* 322–336.

*Roth, C., & Moser, K. (2005). Partizipatives Produktivitätsmanagement (PPM) bei komplexen Dienstleistungen. [The Productivity Measurement and Enhancement System (ProMES; Management system PPM) with complex service tasks.] *Zeitschrift für Personalpsychologie, 4,* 66–74.

Rousseau, D. M. (2005). Is there such a thing as "evidence-based management"? *Academy of Management Review, 31,* 256–269.

Rousseau, D. M. (2006). Keeping an open mind about evidence-based management. *Academy of Management Review, 31,* 1089–1091.

Rousseau, D. M.& McCarthy, S. (2007). Educating managers from an evidence-based perspective. *Academy of Management Learning & Education, 6,* 84–101.

Saavedra, R., Earley, P. C., & van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology, 78,* 61–72.

Salas, E., Kosarzycki, M. P., Tannenbaum, S. I., & Carnegie, D. (2004). Principles and advice for understanding and promoting effective teamwork in organizations. In R. Burke & C. Cooper (Eds.), *Leading in turbulent times: Managing in the new world of work* (pp. 95–120). Malden, MA: Blackwell.

Salas, E., Rosen, M. A., Burke, C. S., Goodwin, G. F., & Fiore, S. (2006). The making of a dream team: When expert teams do best. In K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman, (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 439–453). New York: Cambridge University Press.

Sargent, M. J., & Pritchard, R. D. (2005, April). *The relationship between organizational centralization and productivity improvement.* Poster presented at the 20th annual meeting of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Schmidt, K. H., Kleinbeck, U., & Brockmann, W. (1984). Motivational control of motor performance by goal setting in a dual-task situation. *Psychological Research, 46,* 129–141.

Seijts, G. H., & Latham, G. P. (2001). The effect of distal learning, outcome, and proximal goals on a moderately complex task. *Journal of Organizational Behavior, 22,* 291–307.

Semmer, N., & Frese, M. (1985). Action theory in clinical psychology. In M. Frese & J. Sabini (Eds.), *Goal directed behavior: The concept of action in psychology* (pp. 296–310). Hillsdale, NJ: Erlbaum.

Silverman, S. B., & Wexley, K. N. (1984). Reaction of employees to performance appraisal interviews as a function of their participation in rating scale development. *Personnel Psychology, 37,* 703–710.

Sisk, H. L., & Williams, J. C. (1981). *Management and organization.* Cincinnati, OH: South Western.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for training and simulation* (pp. 61–87). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A., Zeisig, R. L., Action, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271–297). Washington, DC: American Psychological Association.

Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-

analysis, and review of empirical findings. *Personnel Psychology, 58,* 33–66.

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87,* 96–111.

Sturman, M. C. (2003). Searching for the inverted U-shaped relationship between time and performance: Meta-analyses of the experience/performance, tenure/performance, and age/performance relationships. *Journal of Management, 29,* 609–640.

Taylor, M. S., Fisher, D., & Ilgen, D. (1984). Individual reactions to performance feedback in organizations: Control theory perspective. In K. Rowland & G. Ferris (Eds.), *Research in Personnel and Human Resource Management* (pp. 81–124) Greenwich, CT: JAI Press.

Taylor, R., Mulvey, G., Hyman, J., & Bain, P. (2002). Work organization, control, and the experience of work in call centers. *Work, Employment, and Society, 16,* 133–150.

Tubre, T. C., & Collins, J. M. (2000). Jackson and Schuler (1985) revisited: A meta-analysis of the relationships between role ambiguity, role conflict, and job performance, *Journal of Management, 26,*155–169.

Van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20,* 343–352.

Venkatesh, V., & Speier, C. (2000). Creating an effective training environment for enhancing telework. *International Journal of Human–Computer Studies, 52,* 991–1005.

Vroom, V. H. (1964). Work and motivation. New York: Wiley.

Wall, T. D., & Lischeron, J. H. (1977). *Worker participation: A critique of the literature and some fresh evidence.* Maidenhead, England: McGraw-Hill.

Walshe, K., & Rundall, T. G. (2001). Evidence-based management: From theory to practice in health care. *The Milbank Quarterly, 79,* 429–457.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology, 82,* 247–252.

Watrous, K. M., Huffman, A. H., & Pritchard, R. D. (2006). When coworkers and managers quit: The effects of turnover and shared values on performance. *Journal of Business and Psychology, 21,* 103–126.

West, M. A. (1990). The social psychology of innovation in groups. In M. A. West & J. L. Farr (Eds.), *Innovation and creativity at work: Psychological and organizational strategies* (pp. 309–333). Oxford, England: Wiley.

West, M. A. (1994). *Effective team work.* Leicester, England: BPS Books.

West, M. A. (1996). Reflexivity and work group effectiveness: A conceptual integration. In M. A. West (Ed.), *Handbook of work group psychology.* Chichester, England: Wiley.

West, M. A. (2007, May). *Flourishing people teams and organizations: The challenge for work and organizational psychology* Keynote address presented at the European Congress of Work and Organizational Psychology, Stockholm, Sweden.

West, M. A., & Anderson, N. R. (1996). Innovation in top management teams. *Journal of Applied Psychology, 81,* 680–693.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75,* 315–321.

Wood, R. E., & Locke, E. A. (1990). Goal setting and strategy effects on complex tasks. *Research in Organizational Behavior, 12,* 73–109.

Youmans, R. J., & Stone, E. R. (2005). To thy own self be true: Finding the utility of cognitive information feedback. *Journal of Behavioral Decision Making, 18,* 319–341.

Received January 24, 2007
Revision received October 10, 2007
Accepted October 18, 2007 ■