

# The Effect of the Performance Appraisal System on Trust for Management: A Field Quasi-Experiment

Roger C. Mayer  
Baylor University

James H. Davis  
University of Notre Dame

Recent theoretical developments have enabled the empirical study of trust for specific referents in organizations. The authors conducted a 14-month field study of employee trust for top management. A 9-month quasi-experiment found that the implementation of a more acceptable performance appraisal system increased trust for top management. The 3 proposed factors of trustworthiness (ability, benevolence, and integrity) mediated the relationship between perceptions of the appraisal system and trust.

Recognition of the importance of trust in organizations has grown dramatically in recent years, evidenced by an abundance of published work attempting to understand the phenomenon from a variety of perspectives. Trust for employees (Whitney, 1994), trust for management (Mishra, 1996; Robinson, 1996), and interorganizational trust (Gulati, 1995) have all been examined in recent scholarly literature. Many more works have suggested that trust is very important to other phenomena, such as group process and negotiation, without delving into the nature of trust and how it develops.

There is some evidence that suggests that trust levels for management in many organizations are declining (Farnam, 1989). Some researchers have suggested that organizations routinely violate what the employees believe are the employers' obligations, leading to a general erosion of trust for employers (Robinson & Rousseau, 1994). Because trust is so difficult to build (Tyler & Degoe, 1995) and so poorly understood, some authors have suggested legalistic remedies to protect an organization from opportunistic employee behaviors (e.g., Sitkin & Bies, 1994). Argyris (1964) theorized that trust for management is tied to important productivity-related outcomes. Given this, it is important for

both theoretical and practical reasons to understand how to build trust for management.

The effective use of performance appraisal systems may provide an opportunity to build trust in organizations. In the first section that follows, we explore the concept of trust, as well as the factors that are proposed to lead to it. Following that, we consider the relationship between performance appraisal and trust. A quasi-experiment on the effect of a company's performance appraisal system on trust for top management is then presented, followed by a path analysis that examines the role of trustworthiness in mediating the relationship between perceptions of the appraisal system and trust.

## The Concept of Trust

The study of trust has been hampered by its status as a first-degree construct. Calder (1977) has suggested that the study of topics such as trust, which "belong to the world of everyday explanation," (p. 182) leads to a proliferation of approaches to understanding them, because there are numerous connotations of the terms involved. Consequently, although trust has been recognized as important from many organizational perspectives, a number of diverse and somewhat conflicting definitions and approaches have developed.

One example of the approaches that have developed is the sociological approach (e.g., Lewis & Weigert, 1985), in which trust is viewed as a characteristic of the social fabric that facilitates interactions among parties. This approach may be helpful in understanding how a more widespread level of trust among various individuals in a social system can improve the system's ability to function. However, using such an approach, one cannot identify specific actions a particular party might undertake in order to become more

---

Roger C. Mayer, Department of Management, Hankamer School of Business, Baylor University; James H. Davis, Department of Management, College of Business Administration, University of Notre Dame.

We thank Phil Bobko, Kevin Carlson, Ed Conlon, Jeff Edwards, Mark Gavin, Barbara Mayer, Aneil Mishra, Denise Rousseau, and Steve Scullen for their helpful comments.

Correspondence concerning this article should be addressed to Roger C. Mayer, Department of Management, Hankamer School of Business, P.O. Box 98006, Baylor University, Waco, Texas 76798-8006. Electronic mail may be sent to Roger\_Mayer@Baylor.edu.

trusted by a given other or others. Hence, its utility for the current purposes is limited.

A recent framework that appears to be promising as a theoretical foundation for understanding the development of trust defines trust as a willingness to be vulnerable to the actions of another party (Mayer, Davis, & Schoorman, 1995). This conceptualization differentiates trust itself from its outcomes, which are various types of risk-taking in the relationship with the trustee (i.e., to-be-trusted party). Trust defined in this manner does not involve risk *per se*, but is a willingness to engage in risk-taking with the focal party. Such outcomes could include cooperation, sharing sensitive information, and voluntarily allowing the trustee control over issues that are important to the trustor (i.e., trusting party). This conceptualization of trust also separates trust from its antecedents. It holds that a trustor will be willing to be vulnerable to another party based both on the trustor's propensity to trust other people in general, and on the trustor's perception that the particular trustee is trustworthy.

Mayer et al. (1995) posit that trustworthiness is comprised of three factors: ability, benevolence, and integrity. Ability is that group of skills, competencies, and characteristics that allow a party to have influence within some domain. For a member of management, this subsumes both the formal and informal influence they are perceived to have in the organization, as well as their perceived competence and skills. Benevolence is the extent to which a trustee is believed to want to do good to the trustor, aside from an egocentric profit motive. If an employee believes a manager cares about the employee's interests, the manager will be seen as having benevolence for the employee. Integrity is defined as the trustor's perception that the trustee adheres to a set of principles that the trustor finds acceptable. This subsumes not only that a manager espouses values that the employee sees as positive, but also that the manager acts in a way that is consistent with the espoused values.

How these three factors are combined into trustworthiness is idiosyncratic, both between individuals and between situations. In some situations, the trustee's ability may be much more important than the other two factors. Other situations may be composed of tasks that are technically simpler but politically sensitive. In these cases, the trustee's integrity may have a greater impact on trust than does ability. Furthermore, one particular trustor may place a greater amount of weight on one of the factors across various situations than does another trustor.

Mayer et al. (1995) propose that, over time, the trustor will evaluate how positive the outcomes of previous vulnerability are. If the trustor's vulnerability leads to outcomes that the trustor believes are favorable, the trustor will positively reassess some combination of the three factors of the trustee's trustworthiness. Conversely, if the trustor allows vulnerability to the trustee and perceives getting "let down" by the trustee, the trustor will reassess some combi-

nation of trustworthiness factors as being lower than previously believed.

Based on the model, there are at least two ways trust can develop. In addition to evaluation of the outcomes of previous vulnerability to the trustee and a subsequent reevaluation of trustworthiness, the model suggests that other exogenous factors can change the perception of trustworthiness. Because trustworthiness is affected by three distinct factors (ability, benevolence, and integrity), incidents or developments that prompt a reappraisal of any of these perceptions of the trustee will impact trustworthiness. For example, if a trustee's actions reflect movement up a learning curve on some important issue, the perception of the trustee's ability will rise. This movement up the learning curve is not necessarily tied to any prior vulnerability on the part of the trustor, and thus is not caused by the feedback loop from outcomes of vulnerability to trustworthiness.

The model just described (Mayer et al., 1995) considers a party's trust for a specific referent. For the theory's tenets to follow logically, the trustee must be specific, identifiable, and perceived to act with volition. If top management within a company includes an identifiable set of people whose actions have a significant impact on the firm's employees, it makes sense to consider the level of trust which employees have for the top management team. In a small, centralized organization where it is relatively easy to identify who makes what decisions and who has the authority to change organizational policies, the organization and the top decision makers may be seen as equivalent. However, in a larger, more complex organization it is often harder to tell which level of management made what decisions. In such an instance, the organization may take on a life of its own in an employee's eyes as a referent of trust, distinct from top management themselves.

There is evidence that conceptualizing trust as a willingness to be vulnerable has merit. Schoorman, Mayer, and Davis (1996a) studied the delegation of risky tasks by veterinary doctors to hospital staff members. They found that trust for the staff member contributed significantly to the veterinarians' delegation of risky tasks, such as administration of anesthesia by a technician. Along similar lines, in a longitudinal study of a chain of restaurants, Davis, Mayer, and Schoorman (1995) operationalized trust as the employees' willingness to be vulnerable to the facility's general manager. They found that restaurants in which the general managers garnered greater trust from their employees performed significantly better over the following three quarters in sales levels and in net profits, and experienced significantly lower employee turnover rates than restaurants in which the general manager was less trusted.

Taken together, the results of these studies provide evidence that the level of trust for another specific party in an organization affects important processes and outcomes in an organization. Thus, given the apparent utility of viewing

trust as a willingness to be vulnerable in the study of trust in organizational relationships, we used this model in the present study as a basis to investigate the development of trust for top management.

### Performance Appraisal and Trust

Because of the pivotal role that performance appraisal plays in managing human resources (Cardy & Dobbins, 1994), there has been a great deal of research conducted to understand appraisals. Murphy and Cleveland (1991) noted that much of this research has focused on such issues as appraisal formats and minimizing bias from raters. One of the issues that they suggested that warrants further attention is the criteria by which appraisal systems are judged. They suggested that one such research area that needs to be addressed is the reactions of the ratees to the appraisal system. Murphy and Cleveland proposed that research should seek to understand how appraisal accuracy affects reactions to the appraisal. Cardy and Dobbins (1994) mirrored this sentiment and argued that such perceptual reactions to the appraisal system are clearly important to the appraisal system's operational effectiveness.

Along similar lines, Cummings (1983) speculated on the effect that the performance appraisal process can have on employee trust for the organization. He hypothesized that the use of self-appraisal in the performance evaluation system should be positively associated with trust. Further, he proposed that if the results of appraisals are fed back to appraisees, trust will be enhanced.

In light of these researchers' arguments, it makes sense to consider the effect of the performance appraisal system on trust for management. In an organization that has merit-based pay and promotion, perceptions of the performance appraisal system are tantamount. In the performance of their jobs, employees make themselves vulnerable to the organization when they expend effort on their jobs. If an employee expends extra effort in order to reduce product defects or to formulate suggestions about how to improve quality, the employee is dependent upon the appraisal system to detect the increased contribution. If the system fails to be sensitive to the extra contribution, the employee will not receive economic benefits for the labor. This should lower the employee's level of trust for those in the organization responsible for the flawed appraisal system. Conversely, when the appraisal system is seen to clearly reflect the employee's performance, trust for those responsible for the appraisal system should be enhanced.

Murphy and Cleveland (1991) suggested that the acceptability of the performance appraisal system to both raters and ratees is important to the system's effectiveness. They pointed out that acceptance by ratees "is a function of both the process and the outcomes of performance appraisal" (p. 252). On the issue of process, they suggested that both the

extent to which the dimensions being rated are considered relevant to the job and the raters' ability to provide well-informed assessments about performance are important. In the present article this is termed *perceived accuracy*, defined as the extent to which the appraisal system is perceived to accurately tap into relevant behaviors that employees see as contributing value to the organization.

The second acceptability issue in this research, which is due in part to the former issue, is whether the economic outcomes that employees seek are seen as outcomes of higher performance. Critical to the employees is the extent to which the appraisal system allows recognition of and rewards for their contributions. This is referred to in this article as *outcome instrumentality* (Vroom, 1964). The end result of a system wherein the appraisals are not accurate reflections of contributions to the organization is that recognition and rewards that are based on the appraisals would not, in the employees' eyes, be linked with true performance.

One avenue of assessing the impact of a human resource system on trust or trustworthiness is to measure the impact of changing the system on the perceptual and attitudinal variables.

*Hypothesis 1:* Replacement of an appraisal system with one perceived as more acceptable will increase the level of trust for those in management responsible for the change.

The model of trust described earlier in this article theorizes that perceptions of a trustee's ability, benevolence, and integrity form the basis of that party's trustworthiness. The combination of these perceptions can be expected to have a major influence on the level of trust for the trustee. An appraisal system that clarifies and increases the perceived linkage between performance and rewards might be expected to affect all three of the factors of trustworthiness.

The perception of ability (i.e., that group of skills, competencies, and characteristics that allow top management to have influence within some domain) includes the formal and informal influence top management has in the domain of interest. When top management's decisions demonstrate strong competence to the employees, the perception of ability should be affected positively. When management demonstrates skill in understanding and resolving issues that are important to employees, an employee's perception of management's ability should be positively affected. One of the significant roles of management in relation to employees is in defining and measuring performance. An accurate performance appraisal system demonstrates to employees that top management has the important managerial skills needed to manage the organization's workforce. Thus, a performance appraisal system that more accurately measures performance should make top management's capacity to have an impact salient, and should enhance the perception of top management's skills.

*Hypothesis 2:* Replacement of an appraisal system with one perceived as more acceptable will increase the level of perceived ability of those in management responsible for the change.

The perception of top management's benevolence (i.e., the extent to which top management is believed to want to do good to the employee, aside from an egocentric profit motive) should be positively affected by a more accurate appraisal system. A system that more accurately assesses performance so that it can be economically recognized by the company is likely to be received as a signal that top management cares about the employees' interests. An accurate appraisal system that provides developmental feedback to appraisees allows them greater potential for growth and advancement in the company.

In addition to economic benefits such as pay and promotion potential, an accurate appraisal system affords employees a greater opportunity for psychological success (Argyris, 1964), or the feeling that they are successful at what they do. Argyris argued that for an organization to increase an employee's psychological success, it must provide opportunities for the employee to define his or her immediate goals and paths to these goals, relate these to the goals of the organization, and evaluate his or her own effectiveness. We propose here that a trustee's actions that increase the likelihood that the trustor will experience psychological success demonstrate benevolence toward the trustor. Thus, management actions that increase opportunities for an employee's psychological success will increase the employee's perception of management's benevolence.

In an organization that has pay and promotion systems that are based on performance, improving the accuracy of the appraisal system enhances the employee's opportunity to achieve positive outcomes that are both economic and psychological. Management actions that enhance an employee's opportunities for success will be perceived as demonstrating benevolence.

*Hypothesis 3:* Replacement of an appraisal system with one perceived as more acceptable will increase the level of perceived benevolence of those in management responsible for the change.

In a company that espouses a system of performance-based pay and promotions, the belief that performance is accurately measured is critical to management's integrity (i.e., the perception that top management adheres to a set of principles that the employee finds acceptable). Management's use of performance-based reward systems makes it critical that, in the employees' eyes, the system used to evaluate performance is valid. If the appraisal system is faulty, the resulting situation is that management says it will recognize and reward performance, but the actual outcomes will not be consistent with the espoused goals. Those responsible for the policy, who have the capacity to change

the policy, will be considered to have lower integrity because their statements and actions are discrepant. In a situation in which top management communicates their intent to recognize and reward performance, a system that provides a strong linkage between performance and organizational outcomes should positively affect the perception of top management's integrity because their actions more closely reflect their stated values. In common language, it is important that top management "walk the talk": Saying that performance is recognized and rewarded is necessary but not sufficient. The belief that performance and rewards are in fact linked is important to the perception of top management's integrity.

*Hypothesis 4:* Replacement of an appraisal system with one perceived as more acceptable will increase the level of perceived integrity of those in management responsible for the change.

The previous hypotheses suggested that the acceptability of the performance appraisal system impacts not only trust, but each of the three factors of trustworthiness. The model on which this research is based suggests that the three factors of trustworthiness have a direct impact on trust itself. This suggests that the acceptability of the appraisal system may have its impact on trust primarily through its effects on the three trustworthiness factors. In other words, based on the model utilized in this research, the impact of appraisal system acceptability on trust is mediated by the three factors of trustworthiness.

*Hypothesis 5:* The effect of the appraisal system's acceptability on trust for top management is mediated by the perceptions of top management's ability, benevolence, and integrity.

## Method

### *Sample and Procedure*

The data for this study were collected in a small nonunion manufacturing firm in the plastics industry located in a rural area in the Midwest. Over a period of 14 months, three waves of surveys were administered to 166, 170, and 193 production employees and supervisors on company time. This represented .68, .75, and .82 of the company's full-time permanent and temporary production staff in the three respective administrations. The average age of the employees was 36.8 years, and their average tenure with the company was 3.7 years. A total of 45% of the employees were women. In order to ensure confidentiality, the surveys were administered in on-site meetings conducted by a member of the research team. Employees were given time away from their regular duties to complete the surveys.

### *Study Design*

The initial data collection provided a baseline of trust and the trustworthiness factors for top management. Focus group discussions with employees indicated that the performance appraisal

system was not perceived to be acceptable. Specific concerns centered around two issues: The appraisal system did not accurately tap into actual performance, and, as a result, there was a lack of relationship between performance and organizational outcomes. The second survey was administered 5 months after the first. It measured trustworthiness of and trust for top management, as well as employee perceptions of the accuracy and outcome instrumentality of the appraisal system as described below.

Between the second and third waves of surveys, top management implemented a new performance appraisal system. The new system was a standardized, off-the-shelf system selected by top management from a commercial source. The employees were not involved in either its selection or its implementation. The new system contrasted sharply with the old in that it added a self-appraisal on a form parallel to the supervisor's appraisal. In the new system, the supervisor and the employee were required to meet and reconcile their appraisals prior to the supervisor's submission of the evaluation to the company. The major thrust of the new system was to assure that supervisors were discussing specific behaviors and outcomes that were expected of the employees and to make clear that these were the bases of the evaluation. Prior to the third-wave survey, some employees had been appraised under the new system and others had not.

After an initial probationary period, each employee was reviewed annually on the anniversary of the employee's hire date. Thus, an employee's assignment into the "experienced" or "inexperienced" group (i.e., have or have not experienced the new appraisal system firsthand) was based on which month of the year the employee hired into the company. This approximates a random assignment to the conditions, as there is no a priori reason to believe that the month of hire should have a systematic effect on the employees. Analyses of the demographic variables indicated that the two groups did not differ significantly in age, tenure, gender composition, or propensity to trust. There was no single date on which all employees received their raises. Rather, after an employee's appraisal was signed off by management, the employee would receive any raise that was commensurate with the appraisal.

After 9 months had passed, the third-wave survey reassessed the variables measured in the second wave. This allowed for quasi-experimental comparisons of changes over time (i.e., from the second-wave survey to the third-wave survey) between the group that had experienced the new performance appraisal process and the group that had not. Thus, the impact of the performance appraisal system was examined by analyzing changes in the levels of trust of the two groups that had not previously differed across the variables measured.

The quasi-experiment is referred to by Cook and Campbell as an "untreated control group design with pretest measures at more than one time interval" (1979, pp. 117–118). Although they classify an untreated control group design with pre- and postmeasures a "generally interpretable nonequivalent control group design" (p. 103), they state that the design becomes much stronger when additional pretests are added. In their words, "The advantages of pretests at two (or more) time points are considerable" (p. 117). Not only does it allow selection-maturation to be ruled out as an alternative explanation, it also helps to rule out the appearance of a treatment effect because of statistical regression. Although they

support the use of this design because of its interpretability, they posit that the reason it is so seldom used is that it is not usually possible to get access to doing multiple pretests because of the time and costs involved.

## Measures

### Trust Variables

On the basis of the model of organizational trust described earlier in this article, Schoorman, Mayer, and Davis (1996a) developed measures to reflect ability, benevolence, integrity, propensity (based on Rotter, 1967), and trust (10, 12, 13, 7, and 4 items, respectively). The findings indicated that the factors of trustworthiness were distinct in a confirmatory factor analysis and that the scales all had acceptable reliabilities (Cronbach's  $\alpha = .93, .95, .96, .71, .82$ , respectively).

For the present study, the items were altered slightly to reflect a focus on top management. Based on concerns expressed by the company's management about the length of the survey, the measures of ability, benevolence, and integrity were shortened in the current study. For each of these trustworthiness factors, the items that most clearly reflected each theoretical dimension were selected for inclusion. All scales in this study used 5-point Likert-type items with anchors of *agree* and *disagree* for each scale point. All of the previously unpublished scales used in this study can be found in the Appendix.

**Ability.** Six items were used to assess the perception of top management's ability. Scores (Wave 2) ranged from 1.5 to 5. Alphas for this scale were .85 for Wave 2 and .88 for Wave 3.

**Benevolence.** Five items were used to measure the perception of top management's benevolence. Scores ranged from 1.0 to 4.8. Alphas for this scale were .87 and .89 for the second and third waves, respectively.

**Integrity.** The Integrity scale consisted of six items. Scores on the Integrity scale ranged from 1.0 to 4.8. Alphas for this scale were .82 and .88 for waves 2 and 3, respectively.

**Propensity.** The entire eight-item scale used by Schoorman et al. (1996a) was used in this study. Scores ranged from 1.38 to 3.75. Alphas for this scale were .55 and .66 for the second and third waves, respectively.

**Trust.** The four items used by Schoorman et al. (1996a) to measure trust were used in this study. Scores ranged from 1.0 to 5.0. Alphas for this scale were .59 and .60 for the second and third waves, respectively.

### Performance Appraisal Variables

To measure changes in site-specific issues about the performance appraisal that were most salient to the employees, we developed items based closely on employee comments and criticisms made of the system before the change. The measures of perceptions of the performance appraisal system were used as manipulation checks in the quasi-experiment to verify that the issues that were salient to the employees were indeed different under the new system. These measures were also used as exogenous variables to test the mediation hypothesis.

**Perceived accuracy.** The first of the two issues will be referred to as perceived accuracy. This refers to the extent to which the

ratee perceives that the performance rating reflects those behaviors that contribute to the organization. This is similar to accuracy, except that accuracy often carries the connotation that there is some "true" level of performance (Cardy & Dobbins, 1994). Accuracy might be described as the extent to which the rating is consistent with the "true" level of performance. In the current work, the standard against which the employee compares the appraisal is the employee's belief about what matters to the organization. Eight items were included in a scale that assessed the accuracy of the system, defined as the extent to which the appraisal system accurately reflected the ratee's contribution to the organization. These items tapped into such issues as whether the performance evaluation reflected the quantity and quality of work completed. Scores ranged from 1.13 to 4.88. A high score on this scale indicates that the respondent believes that the appraisal system accurately reflects the employee's contribution, whereas a low score indicates the person feels the system does not accurately reflect differences in performance. Alphas for this scale were .82 for the second wave and .70 for the third wave.

**Outcome instrumentality.** The second issue considered under performance appraisal acceptability in this study was the outcome instrumentality of performance. Outcome instrumentality is the belief that performance will lead to desired organizational outcomes. A lack of instrumentality suggests that the employee does not believe that a higher level of actual performance will indeed be followed by improved outcomes such as better pay and improved chances of promotion. Three items were included in a scale that measured outcome instrumentality, defined as the extent to which actual performance was perceived to lead to better pay and promotion potential. Scores ranged from 1.0 to 5.0. A high score on this scale indicates that the employee feels that higher levels of actual performance are rewarded, whereas a low score reflects the respondent's belief that performance and rewards are not related. Alphas for this scale were .72 and .77 for the second and third waves, respectively.

Means, standard deviations, numbers of items, intercorrelations, and Cronbach's alphas of the variables are presented in Table 1.

### Reliability of the Trust Measure

The reliability of the trust measure in this sample deserves further comment. Schoorman et al. (1996a) found a Cronbach's alpha of .82 using the measure of trust used in this study. In the present sample, this measure of reliability was not as high. This raised a concern, as Nunnally (1978) pointed out that the primary effect of random measurement error is that it reduces the likelihood of detecting relationships that actually exist.

Because the levels of trust were not expected to remain stable for the group that experienced the new performance appraisal between Waves 2 and 3, test-retest reliabilities were calculated using data from the group that did not experience the new appraisal system. The test-retest reliability in this sample from Wave 1 to Wave 2 (a 5-month lag) is .75 and from Wave 2 to Wave 3 (a 9-month lag) is .66. Even if trust could be measured in a perfectly reliable way, one should expect some fluctuation in its level over a period of several months as trust grows or declines over time. These test-retest reliability coefficients suggest that even given such expected fluctuations, the scale provided a fairly stable measure of trust over time. On the basis of an analysis of item validities (Kline, 1986), the major reason alpha is low in this sample appears to be the length of the scale. As Cortina (1993) stated, "[alpha] must be interpreted with number of items in mind" (p. 102).

## Results

### Confirmation of the Trustworthiness Factors

Confirmatory factor analyses in LISREL 7 were performed on the trustworthiness factors and trust using listwise deletion. The purpose was to verify that the measures of the three trustworthiness factors tap into constructs distinguishable both from one another and from trust. This was important given that the trustworthiness scales were reduced forms of previously used measures. Table 2 presents the results of these analyses.

Table 1  
*Descriptive Statistics and Intercorrelations*

Variable	M	SD	No. of items	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Ability <sup>a</sup>	3.38	0.65	6	(.85)													
2. Benevolence <sup>a</sup>	2.92	0.74	5	.64**	(.87)												
3. Integrity <sup>a</sup>	3.06	0.67	6	.66**	.69**	(.82)											
4. Propensity <sup>a</sup>	2.65	0.44	8	.19*	.20**	.15*	(.55)										
5. Trust <sup>a</sup>	2.85	0.68	4	.61**	.60**	.58**	.21**	(.59)									
6. Accuracy <sup>a</sup>	3.43	0.69	8	.25**	.28**	.29**	.08	.23**	(.82)								
7. Instrumentality <sup>a</sup>	3.01	0.99	3	.28**	.34**	.40**	.11	.28**	.49**	(.72)							
8. Ability <sup>b</sup>	3.47	0.68	6	.50**	.31**	.37**	.15	.33**	.04	.15	(.88)						
9. Benevolence <sup>b</sup>	3.16	0.76	5	.40**	.43**	.39**	.13	.36**	.15	.34**	.73**	(.89)					
10. Integrity <sup>b</sup>	3.29	0.72	6	.48**	.44**	.50**	.15	.48**	.14	.23*	.74**	.78**	(.88)				
11. Propensity <sup>b</sup>	2.70	0.51	8	.24*	.15	.18	.44**	.17	-.01	-.02	.28**	.27**	.35**	(.66)			
12. Trust <sup>b</sup>	3.04	0.66	4	.40**	.29**	.35**	.12	.55**	.16	.28**	.59**	.65**	.66**	.25**	(.60)		
13. Accuracy <sup>b</sup>	3.48	0.57	8	.14	.13	.21*	.00	.08	.34**	.26**	.22**	.35**	.28**	.04	.25**	(.70)	
14. Instrumentality <sup>b</sup>	3.35	0.94	3	.32**	.23*	.30**	-.08	.08	.16	.39**	.28**	.34**	.33**	.10	.23**	.51**	(.77)

Note. We used a pairwise deletion of cases. Cronbach's alpha values are given in parentheses.

<sup>a</sup> Values from Wave 2. <sup>b</sup> Values from Wave 3.

\*  $p < .05$ . \*\*  $p < .01$ .

Table 2  
*Confirmatory Factor Analyses of Top Management Trust and Trustworthiness Factors*

Model	df	$\chi^2$	GFI	AGFI	RMSR	CFI
Wave 2 ( $n = 166$ )						
Null model	210	1826.30	.23	.16	.321	
Global factor model	189	385.23	.80	.76	.051	.88
Proposed model	183	251.38	.88	.84	.043	.96
Wave 3 ( $n = 185$ )						
Null model	210	2479.36	.18	.10	.364	
Global factor model	189	428.60	.81	.77	.044	.89
Proposed model	183	327.99	.86	.82	.040	.94

*Note.* All models were tested using covariance matrices and maximum likelihood estimation. GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; RMSR = root-mean-square residual; CFI = comparative fit index.

We compared the fit statistics of substantive models with one another. The first might be deemed a global model, in that the items on all four scales were loaded onto a single factor. This model reflects that the respondents do not differentiate trust and the factors of trustworthiness. The second model reflects that the respondents differentiate each factor from the others as the theory proposes.

Examination of the fit statistics in Table 2 reveals that the proposed model provides the best fit with the data in both waves of data in the quasi-experiment. The proposed model fit best on all the fit indexes used: chi-square, goodness-of-fit, adjusted goodness-of-fit, root-mean-square residual, and the comparative fit index. Each of these measures is positively related to a model's fit except the chi-square and root-mean-square, which are inversely related to fit (for descriptions of these fit indexes, see Bentler, 1990; Bentler & Bonnett, 1980; James, Mulaik, & Brett, 1982; Jöreskog & Sörbom, 1989). The comparative fit index (Bentler, 1990) for the proposed model was the only one that exceeded .90. A chi-square difference test was performed to compare the fit of the two substantive models. The difference in chi-squares was significant, for Wave 2,  $\chi^2(6, N = 166) = 133.85, p < .01$ ; for Wave 3,  $\chi^2(6, N = 185) = 100.61, p < .01$ , indicating that the proposed model provided a better fit with the data.

### *Effect on Trust and Trustworthiness*

Between the second and third waves of surveys, top management implemented a new performance appraisal system designed to directly address some of the employees' concerns with the former system (this will be described in more detail below). A *t* test comparing the overall level of trust for top management between the second and third waves of surveys revealed that trust increased a significant, albeit modest, amount during this period,  $t(359) = 2.66, p < .01; d = .28$ . The fact that the data were collected in a single company raises the concern that the change in trust level was due to events other than the implementation of the

performance appraisal system, a threat to internal validity which Cook and Campbell (1979; p. 51) refer to as history. A naturally occurring quasi-experiment allowed the threat of history to be ruled out.

The results of the second-wave survey indicated that employees generally did not believe that the appraisals provided an accurate assessment of their job performance or that their actual performance was related to organizational outcomes. In order to test Hypotheses 1–4, we analyzed data from employees who completed usable surveys in both the second and third waves. Although this necessitated a reduction in the sample size to 79 because of significant ongoing employee turnover and missing data, the quasi-experimental contrast of the extent of change between those who had not been appraised under the new system (i.e., the inexperienced group,  $n = 57$ ) and those who had (i.e., the experienced group,  $n = 22$ ) provides the clearest test of the effect of the new system.

As a manipulation check on whether the implementation of the new system had an effect on appraisal system problems that were salient to the employees, regression analyses were conducted to analyze the changes in accuracy and outcome instrumentality of the appraisal. For each of these two variables, the third-wave assessment was regressed onto both the second-wave assessment and a dichotomous variable representing membership in either the experienced or the inexperienced group. This analysis assessed the impact of the new system while accounting for the prior level of the manipulation check variable. Results of this analysis are presented in Table 3. For both accuracy and outcome instrumentality, the group membership variable was significant,  $t(75) = 2.75, p < .01$ , and  $t(75) = 3.22, p < .01$ , respectively. In both cases, the signs of the regression coefficients were in the expected direction. These results indicate that the new system positively affected the issues about the performance appraisal that were salient to the employees.

As can be seen in Table 4, the inexperienced group

Table 3  
Regression Analyses of Manipulation Checks

Variable	B	$\beta$	$t(75)$	Overall statistics
Accuracy <sup>a</sup>				
Accuracy <sup>b</sup>	.269	.315	2.99**	Multiple $R = .42$
Group membership	.361	.289	2.75**	$R^2 = .17$
				Adjusted $R^2 = .15$
				$F(2, 75) = 7.80^{**}$
Instrumentality <sup>a</sup>				
Instrumentality <sup>b</sup>	.353	.365	3.65**	Multiple $R = .50$
Group membership	.705	.322	3.22**	$R^2 = .25$
				Adjusted $R^2 = .23$
				$F(2, 75) = 12.61^{**}$

<sup>a</sup> Wave 3. <sup>b</sup> Wave 2. \*\*  $p < .01$ .

showed very little increase in trust for top management (mean  $\Delta = .03$ ) compared to the increase of the experienced group (mean  $\Delta = .36$ ). The changes in trust levels for the two groups are depicted in Figure 1. To help rule out alternative explanations based on differences between the two groups,  $t$  tests were conducted to compare the composition of the two groups on trust, ability, benevolence, integrity, propensity, age, and organizational tenure. As can be seen in Table 5, at the outset of the quasi-experiment the two groups did not differ significantly on any of these variables.

Regression analyses were performed to test the effect of the change in appraisal systems. Trust in Wave 3 was regressed onto trust from Wave 2, age, tenure, gender, propensity to trust, and a dichotomous variable indicating which group the person was in (i.e., experienced or inexperienced). This analysis accounts for differences in the de-

Table 4  
Group Comparisons at Wave 2 and Wave 3

Variable	Inexperienced ( $n = 22$ )	Experienced ( $n = 57$ )
Trust		
Wave 2	2.94	2.87
Wave 3	2.97	3.24
Ability		
Wave 2	3.45	3.57
Wave 3	3.37	3.67
Benevolence		
Wave 2	2.95	3.15
Wave 3	3.02	3.39
Integrity		
Wave 2	3.07	3.33
Wave 3	3.12	3.60
Accuracy		
Wave 2	3.56	3.48
Wave 3	3.30	3.57
Instrumentality		
Wave 2	3.05	3.19
Wave 3	2.98	3.64

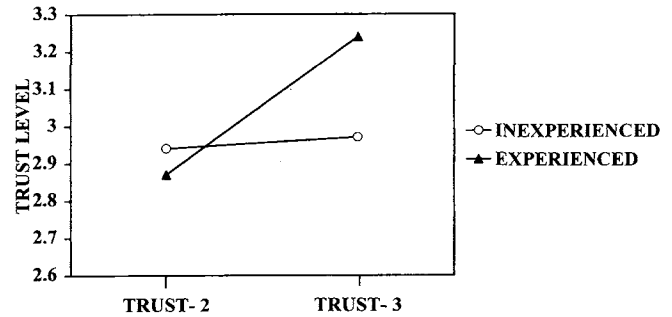


Figure 1. Changes in trust levels for the experienced and inexperienced groups.

mographic variables concurrently with the levels of trust both prior to and after the introduction of the new performance appraisal system. The results of this regression are presented in Table 6. Two of the independent variables were significant predictors of trust in the third wave: trust in the second wave ( $\beta = .60$ ,  $p < .001$ ) and whether or not the person had experienced the new appraisal system ( $\beta = .20$ ,  $p < .05$ ). None of the demographic variables were significant. These results support Hypothesis 1 that a more acceptable appraisal system will lead to a higher level of trust. These results provide stronger support for the assertion that the rise in trust between the second- and third-wave surveys was attributable to the performance appraisal system. This analysis helps to rule out history as an alternative explanation for the increase.

To test Hypotheses 2–4, similar regressions were conducted on ability, benevolence, and integrity. The results of these analyses can also be found in Table 6. In each case, only the prior measure of the particular factor was a significant predictor of that factor in the third wave. Thus, Hypotheses 2–4 were not supported. However, for both benevolence and integrity, the dichotomous variable representing experience with the new appraisal system was close

Table 5  
Preexperimental Group Comparisons

Variable	Inexper.	Exper.	Diff.	$t$	$df$	$d$
Trust	2.94	2.87	.07	0.37	77	.10
Ability	3.45	3.57	.12	0.77	77	.19
Benevolence	2.95	3.15	.20	1.08	77	.27
Integrity	3.07	3.33	.26	1.46	77	.37
Propensity	2.67	2.67	0	0.01	77	0
Accuracy	3.56	3.48	0.08	0.42	76	.12
Instrumentality	3.05	3.19	.14	0.53	76	.14
Age	39.8	38.5	1.4	0.47	74	.11
Tenure	52.1	38.0	14.1	1.17	75	.24
Gender <sup>a</sup>	1.57	1.62	0.05	0.37	75	.10

Note. All variables were measured at Wave 2. Inexper. = inexperienced; Exper. = experienced; Diff. = difference. <sup>a</sup> 1 = female. 2 = male.



Table 6  
Wave 3 Trust and Trustworthiness Regression Analyses

Variable	<i>B</i>	$\beta$	<i>t</i>	Overall statistics
Trust <sup>a</sup>				
Trust <sup>b</sup>	.650	.602	6.55**	Multiple <i>R</i> = .67
Age	.006	.084	0.89	<i>R</i> <sup>2</sup> = .45
Tenure	-.001	-.091	-0.97	Adjusted <i>R</i> <sup>2</sup> = .41
Gender	-.237	-.157	-1.76	<i>F</i> = 9.58**
Propensity	.041	.024	0.26	
Group membership	.324	.196	2.19*	
Ability <sup>a</sup>				
Ability <sup>b</sup>	.623	.538	5.29**	Multiple <i>R</i> = .58
Age	.001	.020	0.19	<i>R</i> <sup>2</sup> = .33
Tenure	-.001	-.078	-0.75	Adjusted <i>R</i> <sup>2</sup> = .27
Gender	.072	.049	0.49	<i>F</i> = 5.71**
Propensity	.079	.047	0.47	
Group membership	.247	.152	1.53	
Benevolence <sup>a</sup>				
Benevolence <sup>b</sup>	.553	.530	5.20**	Multiple <i>R</i> = .58
Age	.004	.056	0.55	<i>R</i> <sup>2</sup> = .34
Tenure	-.001	-.056	-0.55	Adjusted <i>R</i> <sup>2</sup> = .28
Gender	-.066	-.043	-0.44	<i>F</i> = 5.93**
Propensity	.025	.014	0.15	
Group membership	.321	.189	1.91†	
Integrity <sup>a</sup>				
Integrity <sup>b</sup>	.576	.523	5.23**	Multiple <i>R</i> = .61
Age	.005	.071	0.70	<i>R</i> <sup>2</sup> = .37
Tenure	-.001	-.040	-0.40	Adjusted <i>R</i> <sup>2</sup> = .32
Gender	-.114	-.075	-0.77	<i>F</i> = 6.75**
Propensity	.184	.106	1.10	
Group membership	.295	.175	1.78†	

<sup>a</sup> Wave 3. <sup>b</sup> Wave 2. † *p* < .10. \* *p* < .05. \*\* *p* < .01.

to significance (*p* < .07 for benevolence, and *p* < .08 for integrity).

In a quasi-experimental comparison such as this, there is a potential threat to the validity of the comparison between-groups based on the concern that the groups may have already been changing at different rates prior to the time lag under consideration. A follow-up analysis was therefore conducted to determine whether the experienced and inexperienced groups were differentially changing in their levels of trust prior to the second survey. Respondents included in this analysis were those who were both in the quasi-experiment and were also included in the Wave 1 data collection (17 experienced, 41 inexperienced). Trust at Wave 2 was regressed onto both trust at Wave 1 and group membership. Prior trust was a significant predictor of Wave 2 trust,  $t(55) = 7.02$ , *p* < .01, but group membership did not have a significant impact on Wave 2 trust,  $t(55) = -.48$ , *ns*. Thus, the two groups' levels of trust were not changing at different rates in the 5 months prior to the quasi-experiment. This supports the conclusion that the difference between the two groups' changes in trust levels in the quasi-experiment is attributable to the introduction of the new performance appraisal system.

#### Test of Mediation by the Trustworthiness Factors

In order to test Hypothesis 5, that the trustworthiness factors mediate the relationship between appraisal perceptions and trust, LISREL path analyses were conducted using the procedure described by Sapienza and Korsgaard (1996) based on Baron and Kenny (1986). The procedure both compares the fit of alternative substantive models with one another and examines the significance of specific path estimates. The exogenous variables in this analysis were the perceptions of the performance appraisal system. Using a listwise deletion of cases (*n* = 185), the data from Wave 3 were used for this analysis as they reflect the changes in perceptions after the implementation of the new appraisal system. The results of this analysis are presented in Table 7.

Three substantive models were compared with one another: a direct effects model, an indirect effects model, and a saturated model. In the direct effects model we estimated paths from accuracy and outcome instrumentality to ability, benevolence, integrity, and trust. This model represents that accuracy and instrumentality affect trust directly, with no mediation by the trustworthiness variables. The indirect effects model represents the proposed model. It estimates

Table 7  
Model Statistics and Unstandardized Path Coefficients

Measure	Direct	Indirect	Saturated
$\chi^2$	425.17	317.36	316.52
df	6	5	3
Accuracy–Ability	.13	.13	.13
Accuracy–Benevolence	.32**	.32**	.32**
Accuracy–Integrity	.17	.17	.17
Accuracy–Trust	.21*		.06
Instrumentality–Ability	.16**	.16**	.16**
Instrumentality–Benevolence	.18**	.18**	.18**
Instrumentality–Integrity	.20**	.20**	.20**
Instrumentality–Trust	.09		–.03
Ability–Trust		.14**	.15**
Benevolence–Trust		.24**	.23**
Integrity–Trust		.29**	.30**

Note. Data for the LISREL analyses were collected from Wave 3.

\*  $p < .05$ . \*\*  $p < .01$ .

paths from accuracy and instrumentality to ability, benevolence, and integrity, as well as paths from these latter three variables to trust. Both the direct effects model and the indirect effects model are nested under the saturated model. In the saturated model, both direct and indirect effects of accuracy and instrumentality on trust are estimated.

Nested models, which differ in the number of paths estimated, are statistically compared with chi-square difference tests. The difference in chi-squares between the two models is tested for significance, given the difference in degrees of freedom between the two models. A significant difference in chi-squares indicates that the model with fewer degrees of freedom (i.e., the model in which more parameters are estimated) fits better. The first comparison is between the direct effects model and the saturated model to test whether there is any mediation. The second comparison is between the indirect effects model and the saturated

model. This tests whether the effects of accuracy and instrumentality on trust are either fully mediated or partially mediated by the trustworthiness variables. As Sapienza and Korsgaard (1996) noted, the direct effects model and the indirect effects models cannot be compared, because they are not nested.

The comparison between the direct effects model and the saturated model yielded a chi-square of 108.65, with three degrees of freedom ( $p < .001$ ). The significance of this result indicates that the effect of the independent variables on trust was mediated by the trustworthiness variables. The second comparison, between the indirect effects model and the saturated model, yielded a chi-square of .84, with two degrees of freedom (*ns*). The lack of significance of this result indicates that the more complex, saturated model did not improve the fit over that attained by the simpler, indirect effects model.

A follow-up examination of specific path estimates was also conducted. In the saturated model, neither of the paths from the performance appraisal variables to trust were significant. The path from accuracy to benevolence, and the paths from instrumentality to all three of the trustworthiness factors were significant. Likewise, as shown in Figure 2, all three paths from the trustworthiness variables to trust were significant.

The results from both the chi-square tests and the examination of specific paths indicate that the relationship between the performance appraisal variables and trust is fully mediated by the trustworthiness variables. Estimation of direct effects paths from performance appraisal perceptions to trust found that the paths did not improve the fit of the model beyond the fully mediated model, nor were they statistically significant.

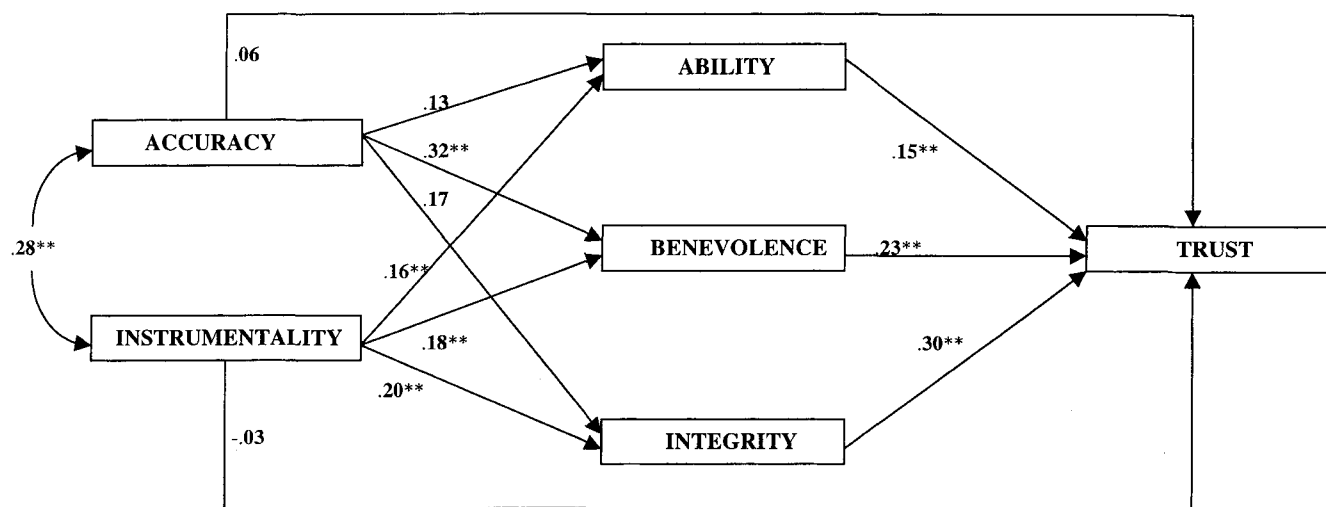


Figure 2. Saturated mediation model of performance appraisal effect on trust. \*\*  $p < .01$ .

## Discussion

Recognition of the importance of trust in organizations has grown considerably in recent years. Tyler and DeGoe (1995) cast doubt on the ability of management to effectively "manage" levels of trust, particularly in the short-term. Although we strongly agree with these authors that attempting to build trust is neither quick nor easy, the results of the current study provide rare evidence that trust might be effectively managed through theoretically based developmental efforts. In this case, those efforts involved replacing a performance appraisal system that was perceived as inaccurate and did not allow for performance-based recognition and rewards. Trust for top management rose significantly in response to the new performance appraisal system. Although history is commonly a threat to internal validity in a study of this nature, one of the strengths of the current study is the quasi-experiment, which helped to minimize this concern. Such field quasi-experiments, particularly in the topic area of trust, are relatively rare. The quasi-experiment described in this study shows that the overall rise in trust for top management was driven by the reactions of the employees who had actual experience with the improved performance appraisal system.

Another strength of this study is the longitudinal nature of the data collection. Data were collected at three points in time over a period of 14 months. This allowed analyses to be conducted that looked at changes in the variables of interest. This within-subjects design and analysis controlled for such factors as variations among individuals' propensity to trust others. Such longitudinal research on the development of trust is relatively rare.

Cardy and Dobbins (1994) discussed the tradeoffs between studying performance appraisal in laboratory and field settings. They appropriately noted that although laboratory studies allow for strict control of the experimental setting, "most laboratory studies have a definite artificial character" (p. 18). They suggest that, in order to investigate such issues as the effects of interventions on outcome measures, the field approach to appraisal research is most appropriate. Although field research cannot by definition provide the extent of control attainable in a laboratory study, the quasi-experiment described in this article was designed and conducted with a great deal of control by field standards. The resulting external validity of the study is very strong. To attempt to study the effect of perceived performance appraisal acceptability on trust for management in a laboratory setting would yield dubious results at best, given the noted concerns with artificiality (Schoorman et al., 1996b).

Along with the strengths of this study, it is also appropriate to consider its limitations. The quasi-experiment described in this study provides strong evidence that those who experienced the new performance appraisal system

showed increased trust for top management, both in an absolute sense and as compared with those who did not experience it. Although this is an important finding, this study does not fully investigate why this is the case. One of the salient improvements in the new system was that its structure necessitated that the rates be given clear, unambiguous feedback about performance and about how to improve it. This was evidenced by the manipulation checks. What cannot be investigated from these data is whether such things as the feedback, ratings, or raises given to the experienced group were higher than whatever had been the case in the past. Top management gave assurances that there had not been a change in policy or practice with respect to raises during the time of the quasi-experiment, but we did not have access to data to empirically test this. Thus, future research should further investigate what aspects of a performance appraisal and reward system have the greatest impact on trust. One framework that would likely be fruitful in extending such investigation would be the differentiation between procedural and distributive justice (Folger, 1987). Perceived accuracy in the current study captures part of what is considered procedural justice, whereas outcome instrumentality in this study is most closely associated with distributive justice.

One potential concern with the current study is that the Cronbach's alpha estimate of reliability of the focal trust measure was lower than that found in previous research. Nunnally (1978) noted that the primary effect of random measurement error is that it reduces the likelihood of detecting relationships that actually exist. If the Cronbach's alpha estimates of reliability reflect random noise in the trust measure, this suggests that the significance of the increase in trust levels may actually underestimate the true significance of the change. Although further developmental work with this measure of trust is warranted, the significance of the rise in trust as predicted, its test-retest reliability (i.e., .75 over 5 months), and its correlations with the other variables (e.g., the correlation with each of the trustworthiness factors was approximately .6) indicate that it provided an adequate measure of the type of trust investigated in this study. Further research on this construct is warranted to investigate whether trust conceptualized as a willingness to be vulnerable consists of multiple dimensions, which would contribute to an attenuation in the level of alpha.

Another potential concern with the current study is what Cook and Campbell (1979) referred to as diffusion of the treatment. Theoretically, we should expect some level of improvement in the perceptions of the appraisal system by those who experience the new system (i.e., the experienced group). They are able to see firsthand exactly what the supervisor is rating, to receive feedback on how they might improve their performance, and to hear an explanation of how their own performance and rewards are related. First-

hand experience with the system as used by their own supervisor is likely to be the most salient evidence that the new system does in fact improve the accuracy of the appraisal and improve the likelihood that their efforts will be rewarded by the organization. In any field setting, one should expect that employees who experience the new system are likely to talk about it with others who have not. If knowledge of the new system were enough to change opinions about the system (and subsequently trust for top management), we should expect to see improvement in the inexperienced (i.e., control) group's perceptions of the appraisal system. If diffusion of treatment affected the results, it should serve to lessen the differences between the two groups (i.e., experienced and inexperienced). To the extent that a diffusion of the treatment occurred, the results obtained would actually understate the effect of the appraisal system.

The company was experiencing substantial voluntary and involuntary turnover as well as absenteeism. Inclusion of a participant in the quasi-experiment necessitated that the person was an employee at both Wave 2 and Wave 3 of the data collection, which were 9 months apart. It also necessitated that the employee was present at work on each of the survey administration days. The employee also had to fill out a useable survey on both days. It is clear from the analyses in this study that trust for top management was differentially affected in the two quasi-experimental groups. What is not known is how the reduction in sample size for the above reasons in the quasi-experiment affected the results. It may be, for instance, that some participants whose trust for top management declined significantly (for whatever reason) prior to the third wave either left the company or failed to respond on the final data collection. Thus, the difference in sample size between a single wave of data and the quasi-experiment represents a limitation to this study.

A somewhat surprising finding in this study was that the new performance appraisal system that was perceived as more accurate did not significantly affect any of the three factors of trustworthiness. Although all of the between-groups (i.e., experienced vs. inexperienced) differences were in the predicted directions, none was by itself significant. Mayer et al. (1995) suggested that trust for a given trustee was determined in part by some combination of the factors of trustworthiness. As noted earlier in this article, the way these factors are combined appears to be complex and idiosyncratic. Although no single factor went up significantly because of the new appraisal system, trust for top management did rise significantly. Furthermore, the impact of the appraisal system's acceptability on trust was fully mediated by the factors of trustworthiness. The data are consistent with the position that the development of trust is affected by an idiosyncratic combination of the three factors of trustworthiness. This pattern of results supports the the-

oretical distinctions among these constructs as proposed by Mayer et al. (1995). How the factors are combined to influence trust warrants further research.

The relatively high correlations between trust and the factors of trustworthiness may give rise to concerns over whether the measure of trust is distinct from the measures of trustworthiness. There is a clear conceptual distinction between trust as a behavioral intention (i.e., willingness to be vulnerable) and perceptions of the trustee (i.e., the factors of trustworthiness). Other empirical analyses in this study support the discriminant validity of trust. First, the confirmatory factor analysis found the best fit with the proposed model (i.e., four factors). The fit statistics for this analysis were clearly above accepted standards (e.g., comparative fit indexes of .96 and .94 in the two waves, respectively). Second, in the quasi-experiment, trust was the only outcome measure that was significantly affected by the new performance appraisal system. If trust were not differentiated from the trustworthiness factors, we should have either found no significant effect on trust, or found the trustworthiness factors to be significantly affected (as was trust). Third, the results of the mediation analysis also support the discriminant validity of trust. If trust were not differentiated from the trustworthiness factors, we should not have found a fully mediated model. If trust were equivalent to the trustworthiness factors, the significant paths from the (exogenous) performance appraisal variables to the trustworthiness variables should have been accompanied by significant paths from the performance appraisal variables to trust. In the saturated model, however, the paths from the exogenous variables to the trustworthiness factors were significant, whereas the direct paths from the exogenous variables to trust were not. In sum, the results of the confirmatory factor analyses, the quasi-experiment, and the mediation analyses all provide evidence of the discriminant validity of trust. These results support the position that trust, viewed as a willingness to be vulnerable, is distinct from the three factors of trustworthiness.

The present study was conducted in a small company of several hundred employees. Many of the production employees had occasional, albeit limited, contact with top management. It is conceivable that this level of contact led study respondents to assess trust differently than they would in a large organization in which contact with top management is less common. Such issues as whether significant company actions are attributed to top management and whether these actions affect the level of trust for top management should also be investigated in a larger organization.

## References

- Argyris, C. (1964). *Integrating the individual and the organization*. New York: Wiley.

- Baron, R. M., & Kenny, D. A. (1986). The mediator-moderator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Calder, B. J. (1977). An attribution theory of leadership. In B. M. Staw & G. R. Salancik (Eds.), *New directions in organizational behavior* (pp. 179-204). Chicago: St. Clair.
- Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Cincinnati, OH: South-Western Publishing.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Cummings, L. L. (1983). Performance-evaluation systems in context of individual trust and commitment. In F. J. Landy, S. Zedrick, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 89-93). Hillsdale, NJ: Erlbaum.
- Davis, J. H., Mayer, R. C., & Schoorman, F. D. (1995, October). *The trusted general manager and firm performance: Empirical evidence of a strategic advantage*. Paper presented at the 15th annual meeting of the Strategic Management Society, Mexico City, Mexico.
- Farnham, A. (1989, December 4). The trust gap. *Fortune*, 120(14), 56-78.
- Folger, R. (1987). Distributive and procedural justice in the workplace. *Social Justice Research*, 1, 143-159.
- Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances. *Academy of Management Journal*, 38, 85-112.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis*. Beverly Hills, CA: Sage.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. London: Methuen.
- Lewis, J., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, 63, 967-985.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709-734.
- Mishra, A. K. (1996). Organizational responses to crisis: The centrality of trust. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 261-287). Thousand Oaks, CA: Sage.
- Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective*. Boston: Allyn and Bacon.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 41, 574.
- Robinson, S. L., & Rousseau, D. M. (1994). Violating the psychological contract: Not the exception but the norm. *Journal of Organizational Behavior*, 15, 245-259.
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651-665.
- Sapienza, H. J., & Korsgaard, M. A. (1996). Procedural justice in entrepreneur-investor relations. *Academy of Management Journal*, 39, 544-574.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (1996a, April). *Empowerment in veterinary clinics: The role of trust in delegation*. Paper presented at the 11th annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Schoorman, F. D., Mayer, R. C., & Davis, J. H. (1996b). Organizational trust: Philosophical perspectives and conceptual definitions. *Academy of Management Review*, 21, 337-340.
- Sitkin, S. B., & Bies, R. J. (1994). The legalization of organizations: A multi-theoretical perspective. In S. B. Sitkin & R. J. Bies (Eds.), *The legalistic organization* (pp. 19-49). Thousand Oaks, CA: Sage.
- Tyler, T., & DeGoey, P. (1995). Trust in organizational authorities: The influence of motive attributions on willingness to accept decisions. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 331-356). Thousand Oaks, CA: Sage.
- Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.
- Whitney, J. O. (1994). *The trust factor: Liberating profits and restoring corporate vitality*. New York: McGraw-Hill.

(Appendix follows)

## Appendix

### Measures of Trust, Trustworthiness, and Performance Appraisal Perceptions

The following instructions prefaced the scales. The anchors shown below were consistent throughout. Headings of construct names are for clarity of exposition, and were not included in the surveys.

Indicate the degree to which you agree with each statement by using the following scale:

1	2	3	4	5
Disagree strongly	Disagree	Neither agree nor disagree	Agree	Agree strongly

Think about [company name]'s top management team [names listed in parentheses for clarity]. For each statement, write the number that best describes how much you agree or disagree with each statement.

#### Ability

- Top management is very capable of performing its job.
- Top management is known to be successful at the things it tries to do.
- Top management has much knowledge about the work that needs done.
- I feel very confident about top management's skills.
- Top management has specialized capabilities that can increase our performance.
- Top management is well qualified.

#### Benevolence

- Top management is very concerned about my welfare.
- My needs and desires are very important to top management.
- Top management would not knowingly do anything to hurt me.
- Top management really looks out for what is important to me.
- Top management will go out of its way to help me.

#### Integrity

- Top management has a strong sense of justice.
- I never have to wonder whether top management will stick to its word.
- Top management tries hard to be fair in dealings with others.
- Top management's actions and behaviors are not very consistent.\*
- I like top management's values.
- Sound principles seem to guide top management's behavior.

#### Propensity

- One should be very cautious with strangers.
- Most experts tell the truth about the limits of their knowledge.
- Most people can be counted on to do what they say they will do.
- These days, you must be alert or someone is likely to take advantage of you.
- Most salespeople are honest in describing their products.
- Most repair people will not overcharge people who are ignorant of their specialty.
- Most people answer public opinion polls honestly.
- Most adults are competent at their jobs.

#### Trust

- If I had my way, I wouldn't let top management have any influence over issues that are important to me.\*
- I would be willing to let top management have complete control over my future in this company.
- I really wish I had a good way to keep an eye on top management.\*
- I would be comfortable giving top management a task or problem which was critical to me, even if I could not monitor their actions.

Think about the performance review system at [company name], and answer the following questions.

#### Accuracy

- The evaluation of what skills I have is pretty accurate.
- How much work I get done is important to my performance review.
- How many mistakes I make in my work is important to my performance review.
- Whether or not my supervisor likes me is important to my performance review.\*
- How much effort I put into my job is important to my performance review.
- How many "extra" things I do is important to my performance review.
- Finding ways for the company to save money is important to my performance review.
- Coming up with good ideas for the company improves my performance review.

#### Outcome instrumentality

- Whether or not I get a raise depends on my performance.
- If you are one of the better performers in this company, you will get one of the better raises.
- If I perform well, my chances of moving up are improved.
- \*-Reverse-scored item.

Received February 27, 1997  
Revision received June 15, 1998  
Accepted June 16, 1998 ■