

Effects of task performance, helping, voice, and organizational loyalty on performance appraisal ratings

By: Steven W. Whiting, Philip M. Podsakoff, [Jason R. Pierce](#)

Whiting, S. W., Podsakoff, P.M., & Pierce, J. R. (2008). The effects of task performance, helping, voice, and organizational loyalty on performance appraisal ratings. *Journal of Applied Psychology*, 93(1), 125-139.

©American Psychological Association, 2008. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: <http://dx.doi.org/10.1037/0021-9010.93.1.125>

***© APA. Reprinted with permission. No further reproduction is authorized without written permission from APA. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. ***

Abstract:

Despite the fact that several studies have investigated the relationship between organizational citizenship behavior and performance appraisal ratings, the vast majority of these studies have been cross-sectional, correlational investigations conducted in organizational settings that do not allow researchers to establish the causal nature of this relationship. To address this lack of knowledge regarding causality, the authors conducted 2 studies designed to investigate the effects of task performance, helping behavior, voice, and organizational loyalty on performance appraisal evaluations. Findings demonstrated that each of these forms of behavior has significant effects on performance evaluation decisions and suggest that additional attention should be directed at both voice and organizational loyalty as important forms of citizenship behavior aimed at the organization.

Keyword: organizational citizenship behaviors | contextual performance | performance appraisals

Article:

Interest in the topic of organizational citizenship behaviors (OCBs) has grown dramatically in the past quarter century (cf. LePine, Erez, & Johnson, 2002; Podsakoff, MacKenzie, Paine, & Bachrach, 2000). Undoubtedly, one of the main reasons for this interest is that OCBs have been shown to enhance organizational effectiveness (cf. Dunlop & Lee, 2004; Koys, 2001; Podsakoff, Ahearne, & MacKenzie, 1997; Podsakoff & MacKenzie, 1994; Walz & Niehoff, 2000). For example, OCBs have been shown to increase both the quantity and quality of work group performance, efficiency, customer satisfaction, profitability, and revenue per full time employee; and decrease customer complaints and employee turnover. Thus, it is probably not surprising that researchers have become interested in the effects that OCBs have on managerial evaluations. If OCBs are related to organizational effectiveness, it is important to know the value that managers place on them in organizational settings. Indeed, in their review of the OCB literature, Podsakoff

et al. (2000) reported on more than a dozen studies that have examined the relationship between OCBs and managerial evaluations of performance (cf. Allen & Rush, 1998; Borman, White, & Dorsey, 1995; Lowery & Krilowicz, 1994; MacKenzie, Podsakoff, & Fetter, 1991; Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo, 1996). Generally speaking, these studies have shown that OCBs account for about as much variance in managerial evaluations of employee performance as task behaviors.

Unfortunately, the vast majority of studies (cf. Allen & Rush, 1998, Study 1; Barksdale & Werner, 2001; MacKenzie et al., 1991; MacKenzie, Podsakoff, & Paine, 1999; Van Dyne & LePine, 1998; Wayne, Shore, & Liden, 1997) that have reported on the relationship between OCBs and managerial evaluations have been cross-sectional, correlational studies that have been conducted in organizational settings. These studies are important because they show that managers do include OCBs in their evaluations in real-life settings. However, they do not allow researchers to establish causal relationships between citizenship behaviors and managerial evaluations. In addition, most of these studies asked managers to provide ratings of both the OCBs and the overall evaluations, thereby raising the possibility that the observed correlations may have been due to same-source or common method biases (cf. Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Williams & Brown, 1994).

One of the few exceptions to the practice of using correlational field data to examine the effects of task performance and OCBs on managerial evaluations is an experimental study reported by Werner (1994). Werner presented information regarding the job performance of several paper people employees to managers who were subsequently asked to evaluate the performance of the employees. The job performance information related to task performance, OCB directed at the organization (OCBO), and organizational citizenship directed at individuals (OCBI). Generally speaking, Werner found that task performance and OCBI had a significant effect on managerial evaluations. However, unlike the previously cited studies, he found that the effect of task performance was substantially greater (accounting for almost 3 times more variance in the managers' evaluations) than that of OCBI.

The study reported by Werner (1994) is important because it established that OCBs do, in fact, influence managerial evaluations of employee performance in a causal sense and that the effect of OCBs on overall performance ratings cannot be attributed solely to same-source biases. This is probably why the study is cited fairly often in the OCB literature. Despite this, several important questions related both to Werner's work in particular and to the causal relationship between OCBs and appraisals in general remain unanswered. Therefore, in this article we report the results of two studies designed to investigate the causal impact of OCBs on appraisal decisions. First we briefly review Werner's study, identify some limitations of his research, and discuss how these limitations affect researchers' understanding of the causal relationship between OCBs and appraisals. Following this, we provide conceptual arguments to explain why three forms of citizenship behavior (helping, voice, and organizational loyalty) should influence managerial appraisals, and we present our hypotheses. Next we report the results of two studies designed to test these hypotheses and thereby enhance investigators' knowledge of the causal relationships between citizenship behaviors and performance evaluations. In the first study, we

investigated the impact of task, helping, and voice behavior on appraisals, whereas in the second study we investigated task, helping, and organizational loyalty. In the final section of the article, we summarize our results and discuss their implications.

Werner's (1994) Study Revisited

For the purposes of his study, Werner (1994) was interested in determining the characteristics that raters considered relevant in appraising the performance of ratees. Specifically, he wanted to examine

the extent to which raters seek out and use information on both role-prescribed and citizenship-type behaviors when making appraisal ratings... [and] to test the extent to which experienced supervisors used both role-prescribed (in-role) ratee characteristics and citizenship (extrarole) characteristics when evaluating performance. (Werner, 1994, p. 98)

Based on his review of the literature, Werner hypothesized that raters would utilize both in-role and extrarole performance information in making their appraisal ratings. In addition, Werner developed hypotheses regarding the information search patterns of raters and the impact that in-role (task performance) and extrarole (citizenship) behaviors would have on raters' halo biases and rating accuracy. However, because we are interested in Werner's results as they relate to the effects that task performance and OCBs have on raters' evaluations, we limit our discussion of his research only to these findings.

In his study, Werner (1994) made critical incidents available to 117 supervisors working in a large state university on an information display board. These incidents related to task performance (job knowledge, and accuracy of work and productivity), OCBO (dependability and attendance, and following policies and procedures) and OCBI (cooperation and teamwork, and extra effort and initiative). Werner found that although raters included both task performance and OCBI in their overall evaluations, task performance accounted for nearly 3 times more variance in their ratings than OCBI. Thus, Werner's findings led to the conclusion that both task performance and OCBs have significant causal effects on managerial evaluations, but that task performance has a substantially greater impact on evaluations than OCBs. However, some limitations of Werner's work impair researchers' ability to interpret these results. Consequently, knowledge regarding the causal relationship between OCBs and performance appraisals is limited by these factors, and much less is known about this relationship than most researchers would conclude.

The first and perhaps most important limitation of Werner's (1994) study is that his manipulation of OCBI was contaminated. One of the dimensions chosen by Werner to represent OCBI was termed *extra effort and initiative* and was defined as "takes on extra tasks when needed and goes the extra mile" (Werner, 1994, p. 100). As such, it has some similarity to task performance or some forms of OCBO such as Graham's (1989) individual initiative construct or Morrison and Phelps's (1999) taking charge construct. Moreover, because exhibiting extra effort and initiative are more likely to benefit the organization at large, rather than specific individuals within it (cf. Williams & Anderson, 1991), it was inappropriate of Werner to include this dimension as a

form of OCBI in his study and doing so raises concerns about the construct validity of his OCBI manipulation.

The second limitation relates to Werner's (1994) treatment of OCBO. To simplify the experimental design, and because Werner felt that OCBOs “straddle the boundary” (p. 101) between task performance and OCBs, he decided not to independently manipulate this form of citizenship behavior in his study. Instead, he yoked the OCBO dimensions to his task performance dimensions. In other words, Werner always assigned the same level of OCBO as the level of task performance that he assigned to each ratee profile. Unfortunately, this makes it impossible to disentangle the independent effects of task behavior and OCBO in the analysis. That is to say, we do not know what portion of the variance attributed to task performance in Werner's study might have been caused by the OCBO manipulations, because these two dimensions were always matched (confounded) with each other. As a result, we do not have a good understanding of the independent effects of task performance, OCBO, and OCBI on rater evaluations in Werner's study.

Third, although Werner (1994) manipulated the task performance and OCBO constructs across three levels (low, medium, and high), he only manipulated the OCBI construct across two levels (medium and high). The problem with this is that the range of values over which task performance and OCBO were manipulated was greater than the range of values over which OCBI was manipulated. This may explain, in part, why task performance was found to have greater effects on raters' evaluations relative to OCBI. Werner noted this weakness in his own discussion, in which he commented that the stronger effect of task behaviors relative to OCBI might simply be the result of an “unfair comparison” (Werner, 1994, p. 105).

One final aspect of Werner's (1994) study that is potentially problematic is that he signaled to participants the type of critical incident that they could observe by labeling each of the six performance dimensions at the top of one of the columns in the 6×6 information display board. Therefore, participants knew what type of information that they would be “sampling” before they actually read the critical incident. We see two potential problems with this procedure. The first is that these labels would play into any “folk theories” (Borman, 1987) that the participants had about the importance of the six dimensions manipulated to overall performance (e.g., that productivity is more important than cooperation in terms of one's overall performance). The second problem with this procedure is that we do not believe that it is reflective of how managers actually sample performance incidents in real-life situations. Indeed, we think it is far less likely that managers in such settings consciously go out and sample a specific form of employee behavior that they have already categorized in their head (e.g., job knowledge and accuracy of work) than it is for them to observe the behavior first, and then assign it to the category in which they feel it belongs. However, regardless of whether we are right or wrong about how managers actually go about this task, we believe that the process of signaling what category the participants in the study are going to be observing is likely to produce different results than when the categories are not signaled and that it is important for researchers to understand these potential differences.

Background and Hypotheses

Within the context of the previous discussion, we had three primary objectives in the studies reported in this paper. First, we set out to examine the impact of OCBI on appraisal decisions, removing the construct contamination that limits Werner's (1994) work. We refer to OCBI as *helping behavior* because this more accurately reflects present-day terminology regarding the subdimensions of OCB, but we note that these terms are essentially interchangeable (Organ, Podsakoff, & MacKenzie, 2006; Podsakoff et al. 2000). Second, we designed these studies to independently manipulate OCBO and thereby enhance the limited knowledge researchers have regarding the impact of these forms of citizenship behaviors on appraisal decisions. Third, in our investigations of the impact of OCBO on appraisals, we sought to update Werner's work from a focus on Organ's (1988) generalized compliance dimension to more contemporary dimensions of citizenship behavior that benefit the organization.

Williams and Anderson (1991) defined OCBO as “behaviors that benefit the organization in general (e.g., gives advance notice when unable to come to work, adheres to informal rules devised to maintain order)... Prior research has labeled... the OCBO dimension as generalized compliance” (pp. 601–602). Werner's (1994) manipulation of OCBO (dependability and attendance, and following policies and procedures) was consistent with this definition and with Organ's (1988) generalized compliance dimension. However, we agree with Werner that generalized compliance tends to straddle the boundary between task behavior and OCB. As a result, we chose not to manipulate generalized compliance in our studies but instead focused on two types of citizenship behavior (voice and organizational loyalty) that (a) are directed at the organization, (b) are of growing interest to researchers in the field (cf. George & Jones, 1997; Graham, 1991; Islam & Zyphur, 2005; Moorman & Blakely, 1995; Morrison & Milliken, 2003; Premeaux & Bedeian, 2003; Van Dyne, Ang, & Botero, 2003; Van Dyne & LePine, 1998), but (c) have not been examined extensively in a performance appraisal context. We outline our rationale for choosing these dimensions, outline their theoretical importance, and develop formal hypotheses in the following sections.

Task Performance

Identifying the critical components of the job performance domain has been a prominent area of investigation for many years (Borman & Motowidlo, 1993; Campbell, 1994; Katz & Kahn, 1978). One obviously important dimension of job performance has been called *in-role behavior* or *task performance*. This category of behavior refers to completing duties and tasks that contribute to the technical core of the organization (Borman & Motowidlo, 1993), that fulfill prescribed job duties and formal job descriptions (Williams & Anderson, 1991), and/or that contribute to the provision of a good or service to customers (Rotundo & Sackett, 2002).

Clearly, one would expect that employee task behaviors should lead to rater judgments regarding performance evaluations. Indeed, it is this very form of behavior that most appraisals are intended to capture and that serves as the basis for rewards that are administered in the formal organizational system. Consistent with this expectation, results from the field have demonstrated that task behaviors are significantly related to appraisal decisions (Borman et al., 1995; Conway, 1999; MacKenzie et al., 1991; Motowidlo & Van Scotter, 1994). In addition, results from lab settings have confirmed that task behaviors lead in a causal sense to increased performance

appraisal ratings (Allen & Rush, 1998; Orr, Sackett, & Mercer, 1989; Rotundo & Sackett, 2002; Werner, 1994). Therefore, we hypothesize the following:

- *Hypothesis 1:* Task performance will have positive effects on performance evaluations such that ratees exhibiting higher levels of task performance will receive higher appraisals.

Helping Behavior

Helping behavior has been defined as “voluntarily helping others with, or preventing the occurrence of work related problems” (Podsakoff et al., 2000, p. 516). At least one form of helping has been included in virtually all conceptualizations of citizenship behavior that have appeared in the literature, including Smith, Organ, and Near's (1983) altruism dimension; Organ's (1988) altruism, courtesy, peacemaking, and cheerleading dimensions; Graham's (1989) interpersonal helping dimension; Williams and Anderson's (1991) OCBI; Borman and Motowidlo's (1993) helping and cooperating dimensions; and Van Scotter and Motowidlo's (1996) interpersonal facilitation dimension.

There are several theoretical reasons to expect that helping will have a significant impact on appraisal ratings that is independent from the impact of task behaviors (MacKenzie et al., 1991; Podsakoff, MacKenzie, & Hui, 1993). One particularly likely explanation of the effects of helping on performance appraisal ratings is the norm of reciprocity. Blau (1964) and Homans (1961), among others, noted that people try to reciprocate a favor to those who have helped them, done them a favor, or treated them fairly. Therefore, when a subordinate performs citizenship behaviors (e.g., by helping a coworker who has fallen behind in his or her work or who has been absent from work because of an illness) that free up the supervisor's time to work on more important matters, one would expect that the supervisor would tend to reciprocate this help by providing a more positive appraisal of the subordinate's performance. Likewise, when an experienced employee provides assistance to a less experienced coworker with a work-related problem that the supervisor would have to deal with otherwise, one would expect the supervisor to reciprocate with a more favorable rating of the experienced employee's performance.

Helping behaviors have also been shown to have a consistent relationship with positive group-level or business-unit outcomes such as the quality and quantity of work produced, customer satisfaction, unit-level profitability, and efficiency (cf. Koys, 2001; Podsakoff et al., 1997; Podsakoff & MacKenzie, 1994; Walz & Niehoff, 2000). As such, helping behavior has direct implications for the bottom line of business units. To the extent that managers explicitly or implicitly understand this relationship, it would make sense for them to consider helping behaviors as an important facet of employee job performance.

Consistent with the above expectations, a substantial amount of cross-sectional correlational evidence has suggested that helping behaviors are positively related to appraisal ratings (cf. Allen & Rush, 1998; Borman et al., 1995; Lowery & Krilowicz, 1994; MacKenzie et al., 1991; Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo, 1996). In addition, what little empirical evidence exists regarding the causal impact of helping behaviors on appraisals has demonstrated that higher levels of helping lead to more favorable appraisal decisions, above and

beyond the impact of task behaviors (Allen & Rush, 1998; D. E. Johnson, Erez, Kiker, & Motowidlo, 2002). Therefore, we hypothesize the following:

- *Hypothesis 2:* Helping behaviors will have positive, independent effects on performance evaluation ratings such that ratees exhibiting higher levels of helping will receive higher appraisals.

Voice Behavior

Van Dyne and LePine (1998) defined employee voice behavior as "... promotive behavior that emphasizes expression of constructive challenge intended to improve rather than merely criticize. Voice is making innovative suggestions for change and recommending modifications to standard procedures even when others disagree" (p. 109). According to the typology developed by Van Dyne, Cummings, and McLean-Parks (1995), voice represents a promotive, challenging form of citizenship behavior. That is, voice challenges the status quo and attempts to promote positive organizational change. By contrast, in this typology, behaviors such as helping are considered promotive, affiliative behaviors that are essentially cooperative or conformist in nature. Therefore, voice behaviors are fundamentally different from other varieties of affiliative OCBs such as helping and loyalty.

There are several reasons why voice behavior might lead to increased appraisal ratings. First, as noted by LePine and Van Dyne (1998; 2001) and Van Dyne and LePine (1998), employee voice might be particularly important in today's business climate. Given the dynamic nature of current work environments, these authors argued that organizations must adapt and change quickly in order to survive and that voice behavior, which challenges the status quo and seeks constructive change, should facilitate the type of change and adaptability required in such dynamic work environments. As a result, managers are likely to view employee voice behaviors as an important component of employee job performance. Second, employees who speak up and provide valuable suggestions or recommendations to managers on how to improve the organization's effectiveness or efficiency would also be expected to help the manager become more successful and subsequently lead him or her to reciprocate these suggestions by providing higher appraisal ratings. Finally, employees who offer suggestions on how to improve the organization are likely to be viewed by their manager as more highly committed to the organization's success than those who choose not to make such suggestions, and because managers' perceptions of employee commitment have been shown to influence performance ratings (Allen & Rush, 1998), this provides another reason why voice behavior would be expected to enhance appraisals.

Interestingly, although a great deal of empirical evidence has converged on the finding that citizenship behaviors are significantly related to performance appraisals, very little of this research has focused on employee voice behavior. Indeed, we are aware of only one study that explicitly investigated the relationship between voice and performance ratings. In this study, Van Dyne and LePine (1998) obtained ratings of task performance, helping, and voice behavior from three different rating sources (supervisors, peers, and self-ratings) and correlated these ratings with supervisor ratings of job performance. They reported that voice behavior was generally positively related to performance appraisal ratings for all three of the rating sources, and both

voice and helping behaviors explained unique variance in managers' evaluations of subordinate performance in at least some of the cases. Despite this generally positive evidence, these authors noted that due to the nature of their data “it is not clear whether helping or voice or some combination of the two contributed to the high supervisory ratings of employee performance” (Van Dyne & LePine, 1998, p. 117). However, considering the arguments discussed earlier, we hypothesize the following:

- *Hypothesis 3:* Voice behaviors will have positive, independent effects on performance evaluation ratings such that ratees exhibiting higher levels of voice will receive higher appraisals.

Organizational Loyalty

Organizational loyalty was defined by Graham (1991) as

identification with and allegiance to organizational leaders and the organization as a whole, transcending the parochial interests of individual, work groups, and departments. Representative behaviors include defending the organization against threats; contributing to its good reputation; and cooperating with others to serve the interests of the whole. (p. 255)

This form of citizenship behavior is obviously different from task performance, and research reported by Van Dyne, Graham, and Dienesch (1994) indicated that it is also distinguishable from other forms of OCB and has somewhat different antecedents and consequences. Related conceptualizations have appeared in the literature in the form of engaging in loyal boosterism (Moorman & Blakely, 1995); spreading goodwill (George & Jones, 1997); and endorsing, supporting, and defending organizational objectives (Borman & Motowidlo, 1993).

One reason why loyalty might be an important determinant of appraisals relates to the common perception that commitment to organizations in present-day work environments is on the decline (L. K. Johnson, 2005). Loyalty behaviors are, in many ways, behavioral manifestations of organizational commitment. Indeed, supporting and defending the organization to outsiders, remaining committed to the organization even in difficult circumstances, and contributing to its good reputation all seem to be behaviors consistent with an attitude of affective commitment to the organization. If managers believe that commitment is on the decline, as reported in the business press (L. K. Johnson, 2005), and if loyalty can be considered a behavioral manifestation of commitment, then it makes sense for raters to value such behaviors in their performance evaluations.

Indeed, beyond the perception that commitment to organizations is declining, there is empirical evidence to suggest that this perception is a reality. For example, when we consulted data from the 1991 and 2002 General Social Survey conducted by the National Opinion Research Center, it appeared that respondents' commitment to their employer declined in this time period. Using a full probability sample of all U.S. households, the General Social Survey asked respondents in 1991 how proud they were of their employing organization (“I am proud to be working for this organization”). In 2002, the General Social Survey asked a similar question of respondents (“I am proud to be working for my employer”). These items were nearly identical to an item

contained on Mowday, Steers, and Porter's (1979) organizational commitment questionnaire. Comparing the means from 1991 to those of 2002 revealed that reports of the amount of pride significantly decreased, $t(2651) = 2.61, p < .01$. Thus, the common perception that commitment is on the decline might accurately reflect the reality of employee attitude change. If so, we believe that it is possible that when raters observe loyalty behaviors, this will be viewed as a manifestation of employee commitment to the organization and these behaviors will be positively valued given their decreasing occurrence in the work place.

Unfortunately, to our knowledge the relationship between organizational loyalty and performance appraisals has not received empirical attention either in field studies or lab research. This is regrettable, given the changing nature of attachment in the workplace and the potential these loyalty behaviors have to influence appraisals as a result. Despite the lack of studies addressing the link between organizational loyalty and performance appraisals, some empirical evidence consistent with the commitment argument is available. For example, in their lab study Allen and Rush (1998) manipulated citizenship behaviors and found that OCB caused rater perceptions of the commitment levels of ratees, and that these commitment ratings subsequently impacted evaluations of ratee performance. Furthermore, in their field study, Allen and Rush demonstrated that commitment partially mediated the relationship between OCB and performance appraisals. Both of these results are consistent with the argument we have made regarding the potential impact of organizational loyalty on appraisals. Indeed, if organizational loyalty behaviors are indicative of commitment to the organization, it makes sense that loyalty will significantly impact rater appraisals of employee performance. Thus, we hypothesize the following:

- *Hypothesis 4:* Loyalty behaviors will have positive, independent effects on performance evaluation ratings, such that ratees exhibiting higher levels of loyalty will receive higher appraisals.

In addition to the hypothesized main effects of task performance and OCBs on performance evaluations, it is possible that these variables interact with one another to influence appraisals. Indeed, previous research has hypothesized and investigated these interactive effects on appraisals and reward decisions (Kiker & Motowidlo, 1999; Rotundo & Sackett, 2002; Werner, 1994). It has typically been proposed that citizenship behaviors will be most important to the appraisals of moderate or high task performing employees, whereas OCBs will have less impact on the appraisals of low task performers. The logic of this argument is that an employee must reach a minimum level of proficiency in task performance before citizenship behaviors will create an impact on appraisals (Kiker & Motowidlo, 1999). However, the empirical evidence reported to date on these interactive effects has been fairly inconsistent (see Organ et al., 2006, for a review); even when significant interaction effects have been found, they have been of relatively small magnitudes. Therefore, we do not formally hypothesize interaction effects in this research. However, we investigate their possible presence in our studies.

Behavioral Labeling

Finally, we also expect the presence or absence of a behavioral category label to interact with task behaviors to influence ratings. Borman (1987) proposed that raters possess implicit performance or folk theories regarding job performance. We believe that one important aspect of such folk theories might be that raters hold an implicit view that in order to be fair and objective in rating performance they should place the greatest emphasis on task-related behaviors, or behaviors that directly address the quantity and quality of work output. These behaviors are likely to be viewed as more important to raters than citizenship-related behaviors. As noted earlier, Werner (1994) provided raters with a category label that signaled to his participants the type of behavior that could be found in each of the columns. Although this is not an unreasonable procedure, we believe that it taps into raters' implicit performance theories and essentially provides a signal to them that they should pay extra attention to the behaviors outlined in the "task behavior" column. As a result, we expect that when participants are told, via a category label, that certain incidents represent task behaviors, they will place greater weight on these incidents in the rating process than they would if the behaviors were not signaled by a label. When behaviors are not signaled, raters must engage in their own categorization process and weight the various behaviors accordingly. Therefore, we hypothesize the following:

- *Hypothesis 5:* Task performance behaviors will significantly interact with behavioral category labels such that the positive relationship between task performance and appraisal ratings will be stronger when category labels are present than when they are absent.

We conducted two studies to investigate these five hypotheses. The purpose of Study 1 was to investigate the causal impact of task, helping, and voice behavior on performance appraisal ratings. The purpose of Study 2 was to examine the causal impact of task, helping behaviors, and organizational loyalty behaviors on appraisal decisions. Finally, in both Studies 1 and 2 we investigated the interaction of task behaviors with behavioral category labels to determine if these labels influenced the importance raters placed on these behaviors.

Study 1

Method

Participants

A total of 137 undergraduate business students from a large public university participated in this study. In exchange for their participation, participants received partial course credit. Their ages ranged from 19 to 36 years, with an average age of 21. Of the participants, 78% were male, and 75% reported that they presently or previously worked full time (30 or more hr per week), with an average combined length of full-time employment of about 14 months. The data for 6 participants were removed from our analysis as these individuals clearly did not complete the experimental task, reading anywhere from no incidents to three total incidents before providing their performance ratings. This resulted in a final sample of 131.

Procedure

Development of the critical incidents

For the purposes of this study, participants were asked to rate the performance of a departmental secretary working in a university setting. As a first step in the study stimulus materials for the performance dimensions of task performance, helping, and voice behavior were developed. These consisted of critical incidents of performance describing the behavior of the target secretary relative to the performance dimension in question. To develop these critical incidents, we first reviewed the original set of incidents used by Werner (1994) to measure task performance and helping behavior. Incidents chosen from this set were supplemented by new incidents created by us. These incidents were written by consulting the formal definitions and measures of the constructs of interest available in the literature, and from interviewing job incumbents. Sample incidents of task, helping, and voice behavior read as follows:

I observed Kim keeping a chart of all she had to do today. This included both verbal and phone requests from the office manager, professors, and graduate students. This probably explains why Kim always completes her assigned tasks on time. [task behavior] Sydney observed Jane (the new departmental secretary) struggling to replace the toner on the copy machine. Even though Sydney was busy at the time, she willingly took time out of her work to show Jane how to properly change the toner. [helping behavior] Terry was not afraid to voice her opinion about how to improve workflow in the department, even though some of the other secretaries thought she should keep quite. [voice behavior]

Given the experimental design of this study, a total of 24 incidents were required. Eight incidents were required for each of the three performance dimensions (task, helping, and voice), four each at two performance levels (high and low performance). This incident generation process resulted in an initial total of 54 incidents to choose from. We reviewed these incidents to determine which were the most construct valid manipulations of task, helping, and voice behavior and to create a logically coherent set of 24 incidents (i.e., eliminating any duplication of incident situations and settings, because participants would have access to the entire set of incidents). This final set of 24 incidents was subjected to a Q-sort, in which nine subject matter experts read and classified each incident. These subject matter experts were doctoral students who had received training in research methods and organizational behavior and who were all familiar with the distinctions between task performance, helping behavior, and voice behavior. They were instructed to indicate which performance dimension each critical incident reflected (task, helping, or voice behavior). Alternatively, if they felt the incident did not represent any of the three performance dimensions provided or that it was relevant to more than one of the three, raters were instructed to classify the incident as "Other." Results indicated adequate levels of agreement across the nine experts. Average agreement for the critical incidents was 81% for task, 78% for helping, and 75% for voice (average $\kappa = .62$).

Study procedures

This study utilized a mixed-subjects design, with three within-subjects factors (task, helping, and voice behavior) each manipulated at two levels (high and low performance) and a single between-subjects factor of header (presence or absence of dimension labels in the header). This resulted in a need for 24 critical incidents of performance related to eight paper people targets. As in Werner's (1994) original design, these targets were arranged in an 8×3 matrix as shown

in Table 1. Participants in the condition where a header was present were provided with the following labels in the column headings: *Task Performance*, *Helping Others at Work*, and *Making Constructive Suggestions*. Participants in the condition where a header was absent were not provided with any column labels. Row and column presentation orders were randomized to eliminate order effects. Additionally, a search constraint was placed on participants, limiting them to viewing 18 behaviors from the matrix (75% of the total behaviors). This mirrored the procedure utilized by Werner and was intended to simulate appraisal situations in real organizations, in which managers cannot conduct a comprehensive information search prior to rating.

Table 1. *Performance Appraisal Means (Study 1)*

Target	Manipulation			Performance appraisal	
	Task	Helping	Voice	<i>M</i>	<i>SD</i>
Lynn	High	High	High	6.04 [5.77]	1.21 [1.55]
Sydney	High	High	Low	5.33 [4.88]	1.34 [1.23]
Kim	High	Low	High	4.92 [4.89]	1.22 [1.55]
Drew	High	Low	Low	4.19 [3.84]	1.10 [1.49]
Pat	Low	High	High	3.74 [4.15]	1.21 [1.32]
Chris	Low	High	Low	3.06 [3.27]	1.51 [1.42]
Terry	Low	Low	High	2.58 [2.55]	1.46 [1.34]
Jody	Low	Low	Low	2.23 [2.57]	1.44 [1.33]

Note. Values outside of brackets are when headers were present ($N = 67$). Values inside of brackets are from when headers were absent ($N = 64$).

Variables

Once participants had finished viewing the critical incidents, they completed a 3-item overall performance appraisal measure adapted from MacKenzie et al. (1991) for each of the eight ratees. A sample item reads “All things considered, this employee is outstanding.” All items used a 7-point Likert scale, anchored at 1 with *strongly disagree* and at 7 with *strongly agree*. The coefficient alpha for the overall performance measure was .97.

Results

To investigate the impact of the experimental factors on overall performance ratings, we utilized a mixed-design analysis of variance with the composite overall performance rating serving as the outcome variable. The three behavioral dimensions (task, helping, and voice) served as within-subjects factors each with two levels (high and low), whereas the header served as a between-subjects factor (presence or absence of a label in the column header). The mean levels of overall performance appraisals assigned to each target for participants both with and without headers are presented in Table 1, and the results of the analysis of variance are shown in Table 2. Additionally, Table 2 provides the partial eta-squared estimates associated with each factor.

Table 2. Performance Appraisal Analysis of Variance Results (Study 1)

Factor	<i>F</i> (1, 129)	Partial η^2
Task performance	340.30**	.73
Helping behavior	156.81**	.55
Voice behavior	70.13**	.35
Header (present vs. absent)	0.05	.00
Task \times Helping	0.03	.00
Task \times Voice	6.58**	.05
Task \times Header	5.63*	.04
Helping \times Voice	3.61	.03
Helping \times Header	0.00	.00
Voice \times Header	0.32	.00
Task \times Helping \times Voice	5.00*	.04
Task \times Helping \times Header	1.37	.01
Task \times Voice \times Header	1.37	.01
Helping \times Voice \times Header	0.62	.00
Task \times Helping \times Voice \times Header	1.35	.01

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

As can be seen in Table 2, task behavior, helping behavior, and voice behavior were all significantly related to performance appraisal decisions (all $F_s \geq 70.13$, all $p_s < .01$). In addition, the effect size estimates reported in this table indicate that task behavior had the highest effect size estimate of .73, followed by helping and voice behaviors with effect size estimates of .55 and .35, respectively. These findings showed that higher levels of task, helping, and voice behaviors led to higher performance appraisal ratings, thus providing support for Hypotheses 1, 2, and 3, respectively. Although the presence or absence of a header did not have a significant main effect on performance appraisal ratings, consistent with Hypothesis 5 column headers did significantly interact with the task performance manipulation to influence performance appraisals, $F(1, 129) = 5.63$, $p < .05$. An examination of the means relevant to this effect indicated that when a column header was present, the relationship between the task performance manipulation and performance appraisals was stronger than when the column header was absent. More specifically, the data showed that when a header was present in the high task performance condition, participants provided slightly higher mean evaluations than when the header was absent ($M = 5.12$ vs. $M = 4.85$), whereas when a header was present in the low task performance condition, participants provided slightly lower mean appraisals than when the header was absent ($M = 2.90$ vs. $M = 3.14$). Thus, this seems to suggest that participants placed greater weight on task behaviors in making their evaluations when column headers were present than when the headers were absent.

A significant two-way interaction between the task performance manipulation and the voice manipulation was also found in this study, $F(1, 129) = 6.58$, $p < .01$; as was a significant three-way interaction effect between the task, helping, and voice manipulations, $F(1, 129) = 5.00$, $p < .05$. The two-way interaction is best understood by an examination of the three-way interaction

effect, which is illustrated in Figure 1. This figure indicates that although helping behavior had little effect on the strength of the relationship between voice behaviors and performance appraisal ratings when task performance was high, the relationship between voice behavior and performance appraisal ratings was stronger when the ratees exhibited high levels of helping when task performance was low. This suggests that raters were less likely to value the suggestions of ratees who performed at low levels unless their low performance was compensated for to some degree by the help that these ratees provided to others.

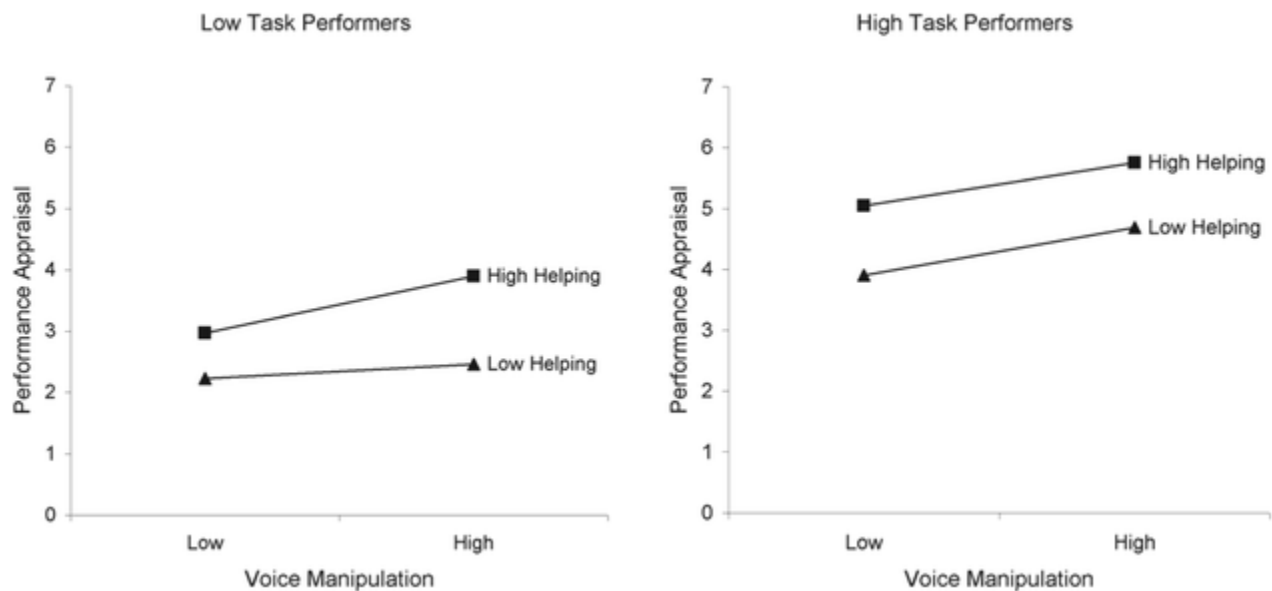


Figure 1. Task \times Helping \times Voice Interaction effect on performance appraisal ratings (Study 1).

Discussion

Consistent with our hypotheses, the results of this study show that manipulations of task, helping, and voice behaviors were each significantly related to performance appraisals. Taken together, these findings indicate that in addition to task behaviors, both helping and voice behaviors have a causal influence on appraisal decisions. The effect size estimates demonstrate that in addition to task behavior, helping and voice behaviors also have a rather substantial impact on appraisal decisions. We also hypothesized that the provision of behavioral labels to raters would impact the relative importance placed by raters on task performance. Results of Study 1 support this hypothesis. Specifically, the significant interaction between the task behavior manipulation and the header manipulation indicates that when participants were told that the behaviors in the task column were task behaviors, they placed greater emphasis on these in making their appraisals than they did when no label was provided. Although we did not directly measure raters' implicit performance theories, this result does provide indirect support for the notion that raters have an implicit theory that in order to provide fair appraisals they should weight task performance more heavily than citizenship behaviors. However, the significant interactive effect has to be tempered by the fact that the mean differences produced by the presence or absence of labels were rather small.

In addition to the hypothesized effects associated with task, helping, and voice behavior on performance appraisals, a significant three-way interaction of these manipulations on appraisals was demonstrated. An examination of these interactive effects indicated that when a target ratee was low on both task and helping behavior, exhibiting high voice behavior had little effect on the appraisal. This finding is interesting, in that it suggests that the value placed on constructive suggestions provided by employees might be tied to their task performance and other citizenship behaviors. Thus, constructive suggestions are given more weight for employees who are perceived to be high in either task or helping behaviors than they are for employees who voice their opinion but are not perceived to contribute much else to the organization. This is noteworthy, as previous thinking regarding the interactive effects of task and OCB on appraisals has used the logic that some minimum level of task proficiency must be reached before citizenship behaviors will have much influence on appraisals. However, this result points to the potential for more complex interactive effects, with particular citizenship behaviors such as voice being dependent on the levels of not only task behavior, but other OCBs as well.

Finally, it should be noted that the pattern of effect sizes demonstrated in this study is rather different from that shown in Werner (1994). As noted earlier, although Werner's results indicated that both task and OCBI (helping) behaviors were significant causal determinants of evaluations, task behavior was much more important than OCBI, with an effect size estimate that was nearly 3 times as large. However, in the present study we found a pattern of effect sizes that was much more representative of results found in previous field studies (Allen & Rush, 1998, Study 1; Barksdale & Werner, 2001; MacKenzie et al., 1991, 1999; Van Dyne & LePine, 1998; Wayne et al., 1997). Thus, although task behavior remains the most important factor in this study, helping and voice both explain substantial amounts of variance in appraisal decisions.

Study 2

The results of Study 1 provided evidence that, in addition to task performance, OCBs directed at other individuals (helping behavior) and at the organization itself (voice behavior) both have a significant effect on the performance appraisal ratings that ratees received. In addition, the findings suggested that the presence or absence of a header did moderate the relationship between task performance and appraisals (although the effect was quite nominal). Building on these findings, the purpose of Study 2 was to investigate the causal impact of task performance, helping behavior, and organizational loyalty behaviors on performance appraisals. As noted earlier, despite the fact that managers are becoming increasingly concerned with organizational loyalty behaviors, research on these types of citizenship behaviors on performance appraisal outcomes is limited, and we are not aware of any studies that have examined this relationship in an experimental setting. Thus, the main focus of Study 2 was to examine the main and interactive effects of task performance, helping behavior, and organizational loyalty on performance appraisals. However, in order to determine whether the findings of Study 1 were replicable, we also again examined whether the presence or absence of the category headers had a moderating effect on the relationship between task performance and performance appraisals.

Method

Participants

A total of 147 undergraduate business students at a large public university participated in this study. In exchange for their participation, participants received partial course credit. Data from 2 participants were removed from this study; one as a result of computer malfunction that rendered the data unusable, and the other because he or she clearly did not take the experimental task seriously, reading only four incidents before providing his or her ratings. This resulted in a final sample size of 145. Participant ages ranged from 19 to 25 years, with an average age of 21. Of the participants, 68% were male, and 72% reported that they either presently or had previously worked full time (30 or more hr per week), with the average combined length of full-time employment being approximately 12 months.

Procedure

Development of critical incidents

This study utilized the same design as that discussed for Study 1, except that the behavioral incidents referring to voice behavior were replaced with incidents relevant to organizational loyalty. To begin to identify the set of stimuli used in this study, we generated an initial sample of 12 critical incidents of loyalty from consulting previously published work and from interviewing job incumbents. This set of 12 incidents was reduced to 8 by identifying which incidents were the most construct valid representations of organizational loyalty and by considering the coherence of the entire set of 24 items (task, helping, and organizational loyalty). A sample loyalty incident reads as follows:

Even when the university had been in the newspaper and received some bad press, Lynn told her friends that she just couldn't believe the story was true, and that the university would be cleared of any wrong-doing in the end.

These eight incidents, along with the incidents from the task and helping dimensions, were submitted to a Q-sort similar to the one performed for Study 1, using nine more subject matter experts whose backgrounds were similar to those of the subject matter experts in the first study. Results of this analysis indicated adequate levels of agreement between the experts. Average agreement for the critical incidents was 89% for task, 90% for helping, and 79% for loyalty (average $\kappa = .83$).

Study procedures

This study used an identical design to that of Study 1, with three within-subjects factors (task, helping, and organizational loyalty) each manipulated at two levels (high and low performance) and a single between-subjects factor of category header. Twenty-four critical incidents of performance were presented for eight paper people targets. Participants in the condition where a header was present were provided with the following labels in the column headings: *Task Performance*, *Helping Others at Work*, and *Loyalty to the Organization*. Participants in the condition where a header was absent were not provided with any labels for the three columns. As in Study 1, the row and column presentation orders were randomized to eliminate order effects.

Additionally, a search constraint was placed on participants, limiting them to viewing 18 behaviors from the matrix (75% of the total behaviors).

Variables

Once participants finished viewing the critical incidents, they completed the same 3-item overall performance appraisal measure used in Study 1. The coefficient alpha of this measure was .97.

Results

The mean levels of overall performance appraisals assigned to each target in both the presence and the absence of the column headers are presented in Table 3, and the results of the analysis of variance are shown in Table 4. Additionally, Table 4 provides the partial eta-squared estimates associated with each experimental factor.

Table 3. *Performance Appraisal Means (Study 2)*

Target	Manipulation			Performance appraisal	
	Task	Helping	Loyalty	M	SD
Lynn	High	High	High	5.68 [5.97]	1.61 [1.27]
Sydney	High	High	Low	5.49 [5.00]	1.20 [1.19]
Kim	High	Low	High	5.07 [5.19]	1.36 [1.45]
Drew	High	Low	Low	4.13 [3.64]	1.51 [1.35]
Pat	Low	High	High	4.02 [4.10]	1.29 [1.53]
Chris	Low	High	Low	2.92 [3.20]	1.32 [1.42]
Terry	Low	Low	High	3.03 [3.20]	1.38 [1.46]
Jody	Low	Low	Low	2.16 [2.09]	1.28 [1.34]

Note. Values outside of brackets are when headers were present ($N = 75$). Values inside of brackets are from when headers were absent ($N = 70$).

Table 4. *Performance Appraisal Analysis of Variance Results (Study 2)*

Factor	$F(1, 129)$	Partial η^2
Task performance	302.42**	0.68
Helping behavior	134.05**	0.48
Voice behavior	182.53**	0.56
Header (present vs. absent)	0.03	0.00
Task \times Helping	0.36	0.00
Task \times Loyalty	0.32	0.00
Task \times Header	1.41	0.01
Helping \times Loyalty	7.26**	0.05
Helping \times Header	0.37	0.00
Loyalty \times Header	6.61**	0.04
Task \times Helping \times Loyalty	6.52**	0.04
Task \times Helping \times Header	0.03	0.00
Task \times Loyalty \times Header	5.51*	0.04
Helping \times Loyalty \times Header	0.35	0.00

Task × Helping × Loyalty × Header	1.47	0.01
-----------------------------------	------	------

As shown in Table 4, task performance, helping behavior, and organizational loyalty behavior were each significantly related to overall performance appraisals (all F s ≥ 134.05 , all p s $< .01$), suggesting that higher levels of these behaviors led to higher performance appraisal ratings. The estimated effect sizes indicated that all three behaviors explained considerable amounts of variance in appraisal decisions. Task behavior had the highest estimate of .68, followed closely by loyalty and helping with estimates of .56 and .48, respectively. These are rather substantial effect sizes, indicating that large amounts of variance in appraisals can be explained by each of the behaviors. These findings were consistent with Hypotheses 1, 2, and 4, respectively. Of interest is the fact that although no support was provided for our expectation (Hypothesis 5) that the presence or absence of a category header would moderate the relationship between task performance and appraisals, the header factor did significantly interact with the loyalty manipulation, $F(1, 143) = 6.61, p < .01$. An examination of the means relevant to this effect indicated that when organizational loyalty was high, overall appraisal scores were lower when a header was present than when one was absent ($M = 4.45$ vs. $M = 4.62$), whereas when loyalty was low, overall appraisal scores were higher when a header was present than when the header was absent ($M = 3.65$ vs. $M = 3.48$). Thus, when a category label was present, participants placed less emphasis on organizational loyalty in making appraisal judgments than when a category label was absent. However, as was the case in Study 1, the mean differences were of a rather small magnitude.

In addition, a significant three-way interaction effect between the task, loyalty, and the header factors was found in this study, $F(1, 143) = 5.51, p < .05$. This interaction effect further informs the effect just discussed, in that it indicated that the header influenced the importance placed on loyalty only in the high task performance condition. Specifically, the impact of loyalty on appraisals was viewed largely the same regardless of a header when task performance was low, but when task performance was high the relationship between organizational loyalty and participants' ratings of ratee performance was somewhat stronger when the header was present than when it was absent.

Finally, a significant two-way interaction between helping and loyalty behavior was demonstrated in this data, $F(1, 143) = 7.26, p < .01$; as was a significant three-way interaction between task, helping, and loyalty, $F(1, 143) = 6.52, p < .01$. These effects are best understood by examining the three-way interaction, which is depicted in Figure 2. As can be seen in this figure, in the case of low task performers the impact of loyalty on evaluations is roughly equivalent for high and low helpers, in that both lines have nearly equal slopes. However, for high task performers, the amount of loyalty exhibited is particularly important to low helpers, in that the slope of the relationship between loyalty and appraisals is somewhat stronger for them than it is for high helpers. Although it is not clear what the explanation is for this somewhat unexpected finding, it suggests that for high task performers, high levels of organizational loyalty may partially compensate for low levels of helping.

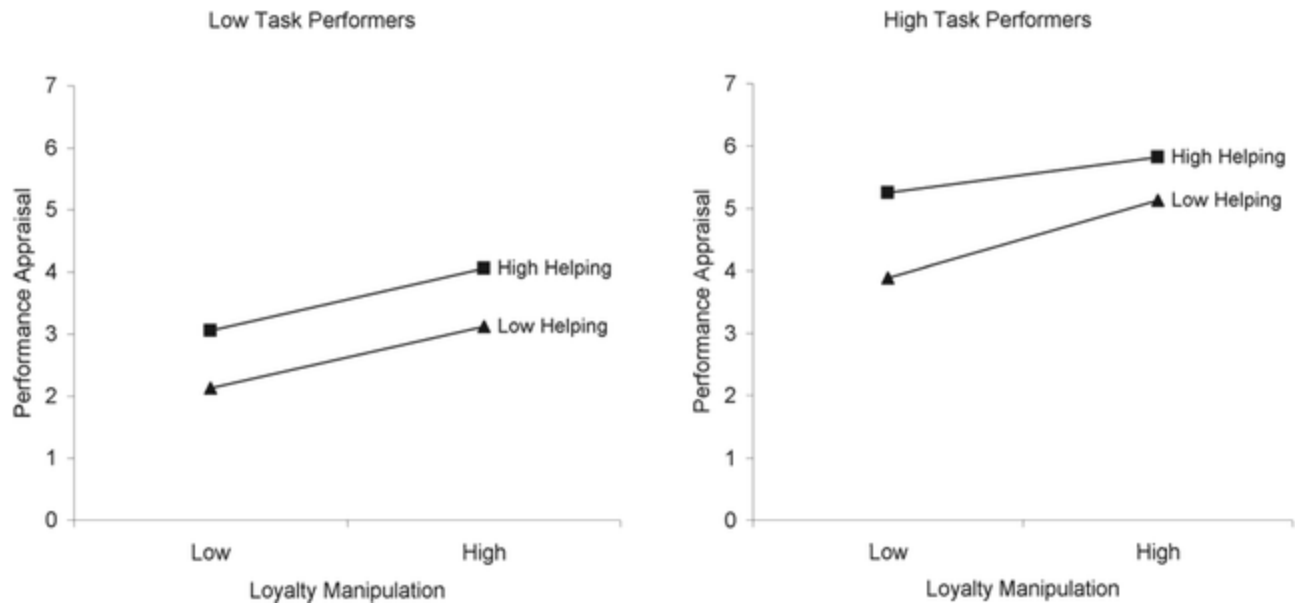


Figure 2. Task \times Helping \times Loyalty interaction effect on performance appraisal ratings (Study 2).

Discussion

Consistent with our hypotheses, we found that task performance, helping behavior, and organizational loyalty behaviors all had significant effects on performance appraisal ratings. Therefore, as in the case of Study 1, these findings indicate that in addition to task performance, OCBs directed at other individuals (helping), as well as those directed at the organization itself (loyalty), are both taken into account in the performance appraisals made by raters. Again, as in Study 1, the citizenship behaviors were not merely trivial causes of appraisals, but rather demonstrated substantial effect sizes that along with task behavior explained a considerable amount of variance in appraisal decisions.

Although we found no support for our expectation that the presence of a header would moderate the effect of task performance on appraisal ratings, we did find a two-way interaction between organizational loyalty and the presence of a header and a three-way interaction between loyalty, task performance, and the presence of a header on appraisal. However, consistent with Study 1, the data from Study 2 show that even though these effects were significant, they had little practical effect on the appraisals.

In addition to the main effects of the three behavioral manipulations on appraisals, a significant three-way interaction between them was found. This interaction indicated that the level of loyalty primarily benefited low helpers in the high task performance condition. More specifically, the relationship between organizational loyalty and appraisals was stronger for low helpers in the high task performance condition than in the low task performance condition. Although this result is somewhat consistent with the argument that employees must reach a minimum level of task performance before OCBs will have much influence on appraisals, it is unclear why it would implicate loyalty in particular. However, this might suggest that when high loyalty is combined

with high task performance it may create a powerful combination that can, in part, compensate for low levels of other behaviors.

Finally, as in the case of Study 1, the pattern of effect sizes demonstrated in this study should be noted here. Werner's (1994) research indicated that task and OCBI (helping) behaviors were both causal determinants of appraisals, but that task behavior was nearly 3 times as important to appraisals as OCBI was. However, the pattern shown here is markedly different. Indeed, although task behavior was shown to be the most influential of the three behaviors on performance appraisals (partial $\eta^2 = .68$), its effects were followed reasonably closely by organizational loyalty (partial $\eta^2 = .56$) and helping behavior (partial $\eta^2 = .48$).

General Discussion

Taken together, these two studies provide strong support for the expectation that, in addition to task performance, helping behavior, voice, and organizational loyalty all contribute significantly to performance appraisal decisions. These findings are important, for several reasons. First, although the finding that helping behavior has a significant effect on performance ratings that is independent of the effect of task performance on these ratings is consistent with previous research, our study is one of the few that has demonstrated this effect in an experimental setting. Therefore, we can have increased confidence that these relationships are not solely the result of the fact that the majority of studies reported in this domain have obtained measures of task performance, OCBs, and performance evaluations from the same source at the same point in time. Thus, we believe that the studies reported herein add to the growing body of knowledge that suggests that helping leads to higher performance appraisal ratings.

Second, this research begins to fill a knowledge gap regarding the causal relationship between voice behavior and appraisal decisions. As noted earlier, we are aware of only one study (Van Dyne & LePine, 1998) that has investigated the relationship between voice behavior and appraisal ratings, and the authors of that study noted that due to the nature of their research it was difficult to determine the extent to which helping, voice, or some combination of the two influenced appraisals. However, the results of Study 1 indicate that voice behaviors are significant causal determinants of appraisal decisions, independent of task performance and/or helping behaviors. This is important, because voice represents a fundamentally different type of citizenship behavior than those investigated in the majority of the field studies reported to date. As noted by Van Dyne and colleagues (1994), voice is a challenging behavior, rather than an affiliative or conformist behavior, and the results of our research indicate that it has a significant positive causal impact on appraisal decisions. Of interest, however, is the fact that our results also indicate that the manner in which raters view voice behavior might be influenced by other subordinate behaviors such as task performance and helping. More specifically, the three-way interaction effect of task, helping, and voice indicated that when both task and helping behaviors were low, a ratee received little benefit from demonstrating high voice behaviors. Given that voice behaviors are challenging rather than affiliative behaviors, this result makes sense in that it suggests that such challenging behaviors might only be valued from individuals who are perceived to be providing other important contributions to the organization.

Third, organizational loyalty was also shown to be a significant causal determinant of appraisals. This is an important finding because very little is known about the relationship between loyalty and appraisal ratings from either experimental or field studies. Indeed, it is fair to say that research on the effects of this form of citizenship behavior is in its infancy. Perhaps, more important, given the changing nature of attachment in present-day organizations (Rousseau & Wade-Benzoni, 1995) and the general view that organizational commitment is on the decline (L. K. Johnson, 2005), there is reason to believe that employee loyalty is of growing importance to managers in organizational settings. Some support for this belief was provided by the somewhat surprising finding of Study 2 that indicated that organizational loyalty had a stronger impact on performance appraisal ratings than did helping behavior. This is surprising, because the research reported to date (cf. Podsakoff et al., 2000) has generally indicated that helping behavior is more strongly related to performance ratings than any other form of OCB. Thus, researchers would be wise to pay additional attention to organizational loyalty and the potential impact that it might have on other decisions such as the manner in which managers allocate pay and promotions in organizational settings.

In addition to the hypothesized effects of task performance and citizenship behaviors on performance appraisal decisions, we also expected that implicit performance theories and/or folk theories would impact the manner in which participants made appraisal decisions. Though we did not directly measure raters' implicit performance theories, the results of our studies do provide some insights into this issue. However, these results are somewhat mixed. Consistent with our expectation, the results of Study 1 indicated that task behavior significantly interacted with the header manipulation, such that participants gave greater weight to task behaviors in making their appraisals when behavioral labels were present than when they were absent. Although the task behavior dimension did not interact by itself with the header manipulation in Study 2, the organizational loyalty dimension did interact with the header, and we found a three-way interaction effect between task performance, organizational loyalty, and the header manipulation. Examination of this interaction effect indicated that when participants were told that the behaviors in a particular column illustrated a ratee's "loyalty to the organization," the participants did not judge these behaviors to be as important to their evaluative decisions as they did when no label was provided. However, this only occurred in the high task performance condition. Despite these interesting findings, it is important to note that in both Study 1 and Study 2 the amount of variance accounted for by these interactive effects was not very substantial. Therefore, even though we found some support for our expectations regarding the moderating effects of providing participants with a header, for all practical purposes the differences produced by this factor were fairly nominal.

Finally, it is important to point out that the pattern of effect sizes we found in Studies 1 and 2 more closely align with results reported in the field than with those reported by Werner (1994). Werner reported that although both task behavior and OCBI (helping) significantly influenced performance appraisals, task behavior was much more important to these appraisals than OCBI, with the partial eta-squared estimate for task behavior being nearly 3 times larger than for OCBI. However, in our studies these differences were much less pronounced. Thus, even though task behavior was still shown to be the most important determinant of appraisals in both Studies 1 and

2 ($\eta^2 = .73$ and $.68$, respectively), helping behavior was also shown to have considerable impact on appraisals ($\eta^2 = .55$ and $.48$, respectively) in these studies. In addition, both voice behavior (Study 1) and organizational loyalty (Study 2) were found to have significant positive effects on raters' appraisal decisions.

These findings are important for several reasons. First, consistent with the arguments of Organ (1988), Borman and Motowidlo (1993), and Rotundo and Sackett (2002), they suggest that raters define job performance more broadly than just task performance, and that citizenship behaviors are included in this broader definition. Second, these findings suggest that raters not only pay attention to citizenship behaviors that are aimed at helping individuals within the organization, but also value citizenship behaviors aimed at helping the organization itself—even when some of these behaviors (e.g., voice behaviors) may be viewed as challenging the status quo. Finally, the findings of our studies suggest that displaying low levels of helping behavior (and to a somewhat lesser extent voice and/or loyalty to the organization) will harm a ratee's appraisal nearly as much as displaying low levels of task behavior. This is important, because it suggests that employees who focus exclusively on the task performance component of their jobs and fail to exhibit any citizenship behaviors are likely to find themselves receiving lower performance evaluations, perhaps leading them to feel that they are treated unfairly because they do not understand the importance that raters place on citizenship behaviors.

Implications for Future Research

Aside from the usual encouragement for researchers to examine the generalizability of our research in other settings, we believe that the findings of the present studies have several important implications for future research. The first is that researchers need to more fully investigate process-oriented explanations for the results shown in this study and in previous field research. Although several theoretical mechanisms, including the norm of reciprocity, informational distinctiveness, implicit performance theories, schema-triggered affect and liking, have been proposed to explain why OCBs would influence managerial evaluations of employee performance (cf. Allen & Rush, 1998; Podsakoff et al. 1993, 2000), it is fair to say that scholars still do not have a very good idea which of these mechanisms have the biggest impact on managerial evaluations. Therefore, we believe that future research designed to investigate these mechanisms would prove to be particularly beneficial to the field.

We also believe that it would be worthwhile to examine the effects that the citizenship behaviors included in the present studies have on other outcome variables. One obvious set of outcome variables that should be included in future research is those that are often associated with managerial evaluations of employee performance, such as reward allocation and promotion decisions. However, we think that it might be particularly interesting to examine the effects of voice and organizational loyalty on group- and/or organization-level outcomes as well. Most of the research (Koys, 2001; Podsakoff et al., 1997; Podsakoff & MacKenzie, 1994; Walz & Niehoff, 2000) that has been conducted at these higher levels of analyses has focused on the relationships between helping behavior and group/organization effectiveness measures. The findings reported to date have been encouraging, in that they have shown that helping behavior is positively related to a variety of effectiveness measures, including product quantity, product

quality, operating efficiency, profitability, and customer satisfaction. However, we are not aware of any research that has examined the potential effects of either voice or organizational loyalty on these types of outcomes, even though there appear to be some good reasons to believe that they should be positively related. Thus, the addition of these forms of citizenship behavior in group-/organization-level research might help improve investigators' understanding of the types of employee behaviors that lead to organizational effectiveness and success.

Finally, we believe that the results of our study indicate that researchers must take greater care in designing experimental studies in this realm to make certain that their manipulations of task performance and OCBs accurately reflect the underlying constructs of interest. Indeed, improving the construct validity of the manipulations, and independently manipulating the behavioral dimensions over similar ranges, resulted in a set of conclusions that are quite different from those reached by Werner (1994). Thus, researchers should take great care regarding such experimental manipulations in the future.

Limitations

Although this article presents several important findings, there are some limitations that should be noted. First, these studies utilized a paper people approach to investigate how various job behaviors impact appraisals. This approach to performance appraisal research has been criticized by some researchers (e.g., Wendelken & Inn, 1981) who believe that paper people methodologies might produce different results than studies using more realistic stimuli. This is consistent with Murphy, Herr, Lockhart, and Maguire's (1986) meta-analysis, which showed a significant difference between the effect sizes of paper people studies and similar studies using direct behavioral observation. However, there are two points worth noting about this criticism. The first is that, although Murphy et al. did report a significant difference between the effect sizes of paper people and observational studies, they found that the percentage of variance accounted for by the type of study (paper people vs. behavioral observation) was quite small ($\omega^2 = .02$). Thus, this statistically significant difference is of little practical importance. The second point is that the results of our studies parallel those of most field studies (cf. Allen & Rush, 1998, Study 1; Barksdale & Werner, 2001; MacKenzie et al., 1991, 1999; Van Dyne & LePine, 1998; Wayne et al., 1997) that show that OCBs account for about as much variance in managerial evaluations as task performance. Thus, although we acknowledge that additional research will be needed to see if our findings can be replicated in other contexts, we do not believe that the use of paper people, in and of itself, represents a significant problem in our research.

Another limitation of this study is the use of student participants. Laboratory research utilizing student participants in performance appraisal contexts has been strongly criticized by some (Guion & Gibson, 1988) and defended by others (Cardy & Dobbins, 1994). However, like Fromkin and Streufert (1976) and Ilgen (1986), we believe the most appropriate way to address this controversy is to try to identify those boundary conditions that might limit the generalizability of laboratory results to field settings based on the use of student participants. In most regards it appears to us, given the task at hand of reading critical incidents of performance and rating the overall performance of paper people targets, that students and practicing managers

would likely provide very similar results. However, it could be argued that practicing managers would have a better understanding of the position in question and would therefore provide different ratings, particularly regarding the importance of the various behaviors to overall appraisals. Although this is of course possible, there are several reasons why we believe this is unlikely to have had a substantial effect on our findings.

First, the best evidence to date (Locke, 1986) regarding the equivalence of laboratory and field studies in organizational behavior research seems to indicate that in multiple topic areas (e.g., staffing, attribution theory, goal setting, participation, financial incentives, job scope–job satisfaction, job satisfaction–job performance), student participants in the lab provide virtually identical or highly similar results to managerial field samples. Second, although a comprehensive analysis of the difference between student and managerial samples with regard to performance appraisal ratings is not available, in Bernardin and Villanova's (1986) analysis of the effects of rater training on performance appraisals, student participants in the lab and managerial participants in the field provided virtually identical results. Thus, although it is possible that practicing managers would provide different results from those shown here with student participants, the best evidence to date regarding the difference between student/laboratory and managerial/field samples in organizational research indicates that this issue is much less problematic than many people believe (Locke, 1986).

A final limitation of this research is that, unlike in Werner (1994), the stimulus materials used for our studies varied across only two levels (high and low) rather than three (high, medium, and low). The potential problem with this strategy is that it may have limited the extent to which certain relationships could emerge. For example, it is possible that exhibiting citizenship behavior might be particularly important for mid-level task performers rather than for high or low task performers. However, because it was our goal to assess the independent effects of task performance, helping, voice, and loyalty in our studies, we were concerned that manipulating these four performance dimensions across three levels would have been too taxing for most participants. Thus, we decided that a simplified design using only two levels of the behavioral factors was appropriate. Although the effects of this limitation should obviously be examined in future research, it is important to note that research on the interactive effects of task behaviors and OCBs has not shown a very consistent pattern of relationships (Allen & Rush, 1998; Kiker & Motowidlo, 1999; MacKenzie et al., 1991; Rotundo & Sackett, 2002), and that virtually none of the studies that have reported significant interactive effects have reported disordinal interactive effects (i.e., ones in which the nature of the relationship between the OCBs and the performance ratings switches signs from positive to negative or vice versa). Therefore, it is not clear that adding an additional level of task, helping, voice, and loyalty behavior would have had any appreciable impact on our results.

Conclusion

Notwithstanding these limitations, the findings of our studies suggest that task, helping, voice, and organizational loyalty behaviors all have significant causal effects on rater performance evaluations. In addition, our findings indicate that although task behaviors have the biggest impact on evaluations, citizenship behaviors in the form of helping, voice, and organizational

loyalty also have substantial effects. Therefore, our studies demonstrate that these forms of behavior are significant and important causal determinants of performance evaluation decisions.

Footnotes

1 We gratefully acknowledge Jon Werner's assistance in providing these incidents.

References

- Allen, T. D., & Rush, M. C. (1998). The effects of organizational citizenship behavior on performance judgments: A field study and a laboratory experiment. *Journal of Applied Psychology, 83*, 247–260.
- Barksdale, K., & Werner, J. M. (2001). Managerial ratings of in-role behaviors, organizational citizenship behaviors, and overall performance: Testing different models of their relationship. *Journal of Business Research, 51*, 145–155.
- Bernardin, H. J., & Villanova, P. (1986). Performance appraisal. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 43–62). Lexington, MA: Lexington Books.
- Blau, P. M. (1964). *Exchange and power in social life*. New York: Wiley.
- Borman, W. C. (1987). Personal constructs, performance schemata, and “folk theories” of subordinate effectiveness: Explorations in an Army officer sample. *Organizational Behavior and Human Decision Processes, 40*, 307–322.
- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 77–98). San Francisco: Jossey-Bass.
- Borman, W. C., White, L. A., & Dorsey, D. W. (1995). Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology, 80*, 168–177.
- Campbell, J. P. (1994). Alternative models of job performance and their implications for selection and classification. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 33–52). Hillsdale, NJ: Erlbaum.
- Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Cincinnati, OH: South-Western.
- Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84*, 3–13.
- Dunlop, P. D., & Lee, K. (2004). Workplace deviance, organizational citizenship behavior, and business unit performance: The bad apples do spoil the whole barrel. *Journal of Organizational Behavior, 25*, 67–80.

- Fromkin, H. L., & Streufert, S. (1976). Laboratory experimentation. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 415–466). Chicago: Rand McNally.
- George, J. M., & Jones, G. R. (1997). Organizational spontaneity in context. *Human Performance*, *10*, 153–170.
- Graham, J. W. (1989). *Organizational citizenship behavior: Construct redefinition, operationalization, and validation*. Unpublished working paper, Loyola University of Chicago, IL.
- Graham, J. W. (1991). An essay on organizational citizenship behavior. *Employee Responsibilities and Rights Journal*, *4*, 249–270.
- Guion, R. M., & Gibson, W. M. (1988). Personnel selection and placement. *Annual Review of Psychology*, *39*, 349–374.
- Homans, G. C. (1961). *Social behavior: Its elementary forms*. New York: Harcourt Brace.
- Ilgel, D. R. (1986). Laboratory research: A question of when, not if. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 257–267). Lexington, MA: Lexington Books.
- Islam, G., & Zyphur, M. J. (2005). Power, voice, and hierarchy: Exploring the antecedents of speaking up in groups. *Group Dynamics: Theory, Research, and Practice*, *9*, 93–103.
- Johnson, D. E., Erez, A., Kiker, D. S., & Motowidlo, S. J. (2002). Liking and attributions of motives as mediators of the relationships between individuals' reputations, helpful behaviors, and raters' reward decisions. *Journal of Applied Psychology*, *87*, 808–815.
- Johnson, L. K. (2005). The new loyalty: Make it work for your company. *Harvard Management Update*, *10*(3), 3–5.
- Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York: Wiley.
- Kiker, D. S., & Motowidlo, S. J. (1999). Main and interaction effects of task and contextual performance on supervisory reward decisions. *Journal of Applied Psychology*, *84*, 602–609.
- Koys, D. J. (2001). The effects of employee satisfaction, organizational citizenship behavior, and turnover on organizational effectiveness: A unit-level, longitudinal study. *Personnel Psychology*, *54*(1), 101–114.
- LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, *87*, 53–65.
- LePine, J. A., & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of Applied Psychology*, *83*, 853–868.

- LePine, J. A., & Van Dyne, L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential relationships with Big Five personality characteristics and cognitive ability. *Journal of Applied Psychology, 86*, 326–336.
- Locke, E. A. (Ed.). (1986). *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management*. Lexington, MA: Lexington Books.
- Lowery, C. M., & Krilowicz, T. J. (1994). Relationships among nontask behaviors, rated performance, and objective performance measures. *Psychological Reports, 74*, 571–578.
- MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons performance. *Organizational Behavior and Human Decision Processes, 50*, 123–150.
- MacKenzie, S. B., Podsakoff, P. M., & Paine, J. E. (1999). Do citizenship behaviors matter more for managers than for salespeople? *Journal of the Academy of Marketing Science, 27*, 396–410.
- Moorman, R. H., & Blakely, G. L. (1995). Individualism-collectivism as an individual difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior, 16*, 127–142.
- Morrison, E. W., & Milliken, F. J. (2003). Speaking up, remaining silent: The dynamics of voice and silence in organizations. *Journal of Management Studies, 40*, 1353–1358.
- Morrison, E. W., & Phelps, C. C. (1999). Taking charge at work: Extrarole efforts to initiate workplace change. *Academy of Management Journal, 42*, 403–419.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475–480.
- Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior, 14*, 43–77.
- Murphy, K. R., Herr, B. M., Lockhart, M. C., & Maguire, E. (1986). Evaluating the performance of paper people. *Journal of Applied Psychology, 71*, 654–661.
- Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
- Organ, D. W., Podsakoff, P. M., & MacKenzie, S. B. (2006). *Organizational citizenship behavior: Its nature, antecedents, and consequences*. Thousand Oaks: Sage.
- Orr, J. M., Sackett, P. R., & Mercer, M. (1989). The role of prescribed and nonprescribed behaviors in estimating the dollar value of performance. *Journal of Applied Psychology, 74*, 34–40.
- Podsakoff, P. M., Ahearne, M., & MacKenzie, S. B. (1997). Organizational citizenship behavior and the quantity and quality of work group performance. *Journal of Applied Psychology, 82*, 262–270.

- Podsakoff, P. M., & MacKenzie, S. B. (1994). Organizational citizenship behaviors and sales unit effectiveness. *Journal of Marketing Research*, 3, 351–363.
- Podsakoff, P. M., MacKenzie, S. B., & Hui, C. (1993). Organizational citizenship behaviors as determinants of managerial evaluations of employee performance: A review and suggestions for future research. In G. R. Ferris & K. M. Rowland (Eds.), *Research in personnel and human resources management (Vol. 11)*, pp. 1–40. Greenwich, CT: JAI Press.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903.
- Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, 26, 513–563.
- Premeaux, S. F., & Bedeian, A. G. (2003). Breaking the silence: The moderating effects of self-monitoring in predicting speaking up in the workplace. *Journal of Management Studies*, 40, 1537–1562.
- Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, 87, 66–80.
- Rousseau, D. M., & Wade-Benzoni, K. A. (1995). Changing individual-organization attachments: A two-way street. In A. Howard (Ed.), *The changing nature of work* (pp. 290–322). San Francisco: Jossey-Bass.
- Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, 68, 655–663.
- Van Dyne, L., Ang, S., & Botero, I. C. (2003). Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of Management Studies*, 40, 1359–1392.
- Van Dyne, L., Cummings, L. L., & McLean-Parks, J. M. (1995). Extra-role behaviors: In pursuit of construct and definitional clarity (a bridge over muddied waters). In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior (Vol. 17)*, pp. 215–285. Greenwich, CT: JAI Press.
- Van Dyne, L., Graham, J. W., & Dienesch, R. M. (1994). Organizational citizenship behavior: Construct redefinition, measurement and validation. *Academy of Management Journal*, 37, 765–802.
- Van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: Evidence of construct and predictive validity. *Academy of Management Journal*, 41, 108–119.
- Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, 81, 525–531.

- Walz, S. M., & Niehoff, B. P. (2000). Organizational citizenship behaviors: Their relationship to organizational effectiveness. *Journal of Hospitality & Tourism Research*, *24*, 301–319.
- Wayne, S. J., Shore, L. M., & Liden, R. C. (1997). Perceived organizational support and leader-member exchange: A social exchange perspective. *Academy of Management Journal*, *40*, 82–111.
- Wendelken, D., & Inn, A. (1981). Nonperformance influences on performance evaluations: A laboratory phenomenon? *Journal of Applied Psychology*, *66*, 149–158.
- Werner, J. M. (1994). Dimensions that make a difference: Examining the impact of in-role and extrarole behaviors on supervisory ratings. *Journal of Applied Psychology*, *79*, 98–107.
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, *17*, 601–617.
- Williams, L. J., & Brown, B. K. (1994). Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes*, *57*, 185–209.