The Effects of Citizenship Performance, Task Performance, and Rating Format on Performance Judgments

by

David R. Coole

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Walter Borman, Ph.D.
Carnot Nelson, Ph.D.
Marcia Finkelstein, Ph.D.

Date of Approval: October 24, 2003

Keywords: OCB, assessment, evaluation, appraisal, feedback

© Copyright 2003, David R. Coole
List of Tables

Table 1  Coleman and Borman’s Taxonomy of Citizenship Performance  17
Table 2  Summary Table for Overall Performance Ratings  25
Table 3  Summary Table for Task Performance Ratings  27
List of Figures

Figure 1. The hypothesized interaction between citizenship and task performance levels as a function of overall performance ratings. 12

Figure 2. The hypothesized interaction between citizenship performance level and rating format as a function of task performance ratings. 14

Figure 3. Overall performance ratings as a function of citizenship and task performance levels. 24

Figure 4. The effects of citizenship performance level on ratings of task performance as a function of rating format. 26

Figure 5. The effects of citizenship performance level on ratings of task performance pooled across rating formats. 28
The Effects of Citizenship Performance, Task Performance, and Rating Format on Performance Judgments

David R. Coole

ABSTRACT

The current study examined the effects of citizenship performance, task performance, and rating format on overall and task performance ratings. Levels of citizenship performance (high, medium, low), task performance (high, medium, low), and rating format (inclusion or exclusion of citizenship performance) were experimentally manipulated in a 3x3x2 between-subjects full factorial design. Ratings were provided by 360 undergraduate psychology students evaluating experimentally developed supervisory logs of first line financial managers. Targets’ levels of citizenship and task performance were positively related to raters’ judgments of overall and task performance. The prediction that this relationship would be moderated by task performance level was not supported. Furthermore, replicating the findings of J. M. Werner (1994), task performance ratings, assigned to targets with high levels of citizenship performance, displayed significantly more halo than ratings assigned to targets with low or medium levels of citizenship performance. Rating format did not influence raters’ judgments of the targets’ overall or task performance. Our findings indicate that including OCBs in job performance assessment fails to increase the accuracy of performance ratings. Study implications and limitations are discussed.
Introduction

In recent years, organizational citizenship behavior (OCB) has become one of the most popular areas of interest for industrial organizational psychologists. Since its introduction to the I/O literature in 1983 (Smith, Organ & Near, 1983), research has explored the construct domain of OCB and has linked OCB with organizational effectiveness, overall employee performance, Big Five personality traits, work attitudes, and procedural justice (Podsakoff, MacKenzie, Paine, & Bachrach, 2000). These research findings have changed the way performance is defined, assessed, and compensated.

Though OCB has had substantial effects on human resource activities including selection, training, and performance evaluation, more research is needed to understand the OCB construct domain and how it interacts with other dimensions and predictors of job performance. The current study investigates the relationship between citizenship performance, task performance, and overall performance. More specifically, the study examines the interactive effects of citizenship performance, task performance, and rating format on performance ratings.

Construct Definition and Development

Organ popularized the construct of OCB as “individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization” (1988, p. 4).
Since then, there has been much debate concerning the extent to which OCBs are in-role or extra-role behaviors, and whether or not they are recognized by reward systems.

Contributing to the debate, Borman & Motowidlo (1993) redefined OCB by introducing *contextual performance* to the literature. In doing so, they combined elements of Organ’s conceptualization of OCB, Brief and Motowidlo’s (1986) theory of prosocial organizational behavior, and the model of soldier effectiveness set forth by Borman and his colleagues (1983). Using facets across these three related topics, they defined contextual performance as behaviors that shape “the organizational, social, and psychological context that serve as a catalyst for task activities and processes” (p. 71). As prescribed by their definition, contextual performance is not bound to extra-role behaviors nor is it beyond the recognition of reward systems. Borman and Motowidlo’s taxonomy of contextual performance consisted of five dimensions: (a) persisting with enthusiasm and extra effort as necessary to complete own task activities successfully; (b) volunteering to carry out task activities that are not formally part of own job; (c) helping and cooperating with others; (d) following organizational rules and procedures; and (e) endorsing, supporting, and defending organizational objectives.

Due to the similarities of these dimensions to the original construct domain of OCB including altruism, compliance, courtesy, sportsmanship, and civic virtue, Organ has reevaluated his original definition of OCB (Organ, 1997). The revised construct now includes in-role and extra-role behaviors that may or may not be recognized by organizational reward systems. Organ and Paine (1999, p. 4) have argued that OCBs may not be a part of the formal job description, but there will often be “expectations by
peers, bosses, or the individual that those aspects of performance be rendered.” In light of these concessions, most researchers and applied psychologists semantically interchange OCB and contextual performance.

More recently, researchers have attempted to summarize and configure the numerous dimension sets of contextual performance by renaming the performance domain as *citizenship performance* and settling on a representative dimensional model incorporating several related dimension sets (e.g. OCB, extra-role behavior, prosocial organizational behavior, etc.), (Coleman & Borman, 2000). Coleman and Borman attempted to identify a single citizenship dimension set representing behavioral dimensions from 14 OCB related studies. The researchers had 47 industrial-organizational psychologists sort behavioral examples representing 27 dimensions into categories based on content similarity. Using the methodology employed by Borman and Brush (1993), the sorting task allowed for the development of a pooled similarity matrix and the derivation of an indirect similarity matrix. Factor analyses, multidimensional scaling analyses, and cluster analyses of the derived matrix resulted in the emergence of a consistent three-factor solution. These factors were: (a) personal support, (b) organizational support, and (c) conscientious initiative.

Borman and Motowidlo’s five-factor model of contextual performance fits neatly into the three-factor representation of citizenship performance. Personal support represents the single dimension of helping others; organizational support combines the original facets reflecting conscientiousness and organizational loyalty; conscientious initiative pools the elements of volunteering and extra effort. The parsimony and stability
of the three-factor model has made it the central construct paradigm for recent research efforts by Borman and his colleagues (Borman et. al., 2001). For the same reasons, the three-factor model of citizenship performance is utilized in the current study.

**Antecedents of OCB**

Most of the early research on OCB examined its antecedents or predictive correlates. Numerous studies have focused on the link between OCBs and dispositional or attitudinal predictors. The most comprehensive review pertaining to the dispositional and attitudinal predictors of OCB has been Organ and Ryan’s meta-analysis (1995). They reviewed 55 studies investigating the correlates of OCB. Correcting for unreliability, these researchers found weighted mean correlations demonstrating a link between job attitudes and facets of OCB. Following the meta-analytical procedures of Hunter and Schmidt (1990), Organ and Ryan found OCB dimensions of general compliance and altruism to be positively related to employee attitudes reflecting job satisfaction (mean corrected r’s = .28, .28), leadership consideration (mean corrected r’s = .35, .32), organizational commitment (mean corrected r’s = .32, .25), and perceptions of organizational justice (mean corrected r’s = .27, .24). Although the causal direction of the relationship between OCB and job satisfaction is still a topic of debate (Organ & Paine, 1999), it is clear that OCBs are more likely to be displayed if employees like their jobs, have intentions of staying at their jobs, and feel they are being treated fairly. Supportive and considerate leaders are likely to further facilitate the display of OCB.

Organ and Ryan’s meta-analysis has also provided evidence for relationships between personality characteristics and OCBs. These authors investigated the
relationships between altruism and general compliance and four personality traits, conscientiousness, agreeableness, positive and negative affectivity. They found corrected mean correlations between both OCB dimensions and conscientiousness (mean corrected r’s = .22 and .30, respectively). The researchers also reported weaker corrected correlations between agreeableness and both OCB dimensions (mean corrected r’s = .13 and .11, respectively), and between positive affectivity and altruism (mean corrected r = .15). Correlations between OCB dimensions and both conscientiousness and agreeableness were lower when analyses controlled for self-report measures of OCB (mean corrected r’s = .04 -.23).

Although early research reported modest correlations between OCB dimensions and personality traits, more recently, researchers have reported stronger relationships. Miller, Griffin, and Hart (1999) found correlations as high as .42 between conscientiousness and citizenship performance even after neuroticism and extraversion had been controlled. Neuman and Kickul (1998) found correlations ranging from .20 and .41 between conscientiousness and all five of Organ’s original components of OCB. Tillman (1998) found conscientiousness to correlate .55 with a composite measure of OCB in a sample of working college students.

Though not as robust, support for personality constructs other than conscientiousness has been reported in research spanning recent years. The Borman, Penner, Allen and Motowidlo (2001) review of personality and citizenship performance yielded mean uncorrected correlations between several personality traits and OCB. Collapsing across OCB dimensions they found mean correlations ranging from .13 to .28
between OCB and personality characteristics including: conscientiousness (mean $r = .24$), agreeableness (mean $r = .13$), positive affectivity (mean $r = .18$), negative affectivity (mean $r = -.14$), locus of control (mean $r = .16$), collectivism (mean $r = .15$), and prosocial personality (mean $r$’s = .22 - .28). These correlations were representative of 20 studies published subsequent to the Organ and Ryan review. Considering the findings of both reviews, there is moderate support for a positive link between conscientiousness and OCB. Research investigating the relationship between other personality characteristics and OCB has been mixed but promising. The research to date has made a reasonably strong case for the dispositional nature of OCB.

Relationships Between OCB and Organizational Effectiveness

Understanding the antecedents of OCB has little value unless its construct domain is related to the successful functioning of an organization. Recognizing this, Organ (1988) hypothesized that OCB is linked to organizational performance. Podsakoff and MacKenzie (1997, p. 138) have postulated that OCBs have several channels through which they create advantageous outcomes for organizations:

a) Enhancing coworker and managerial productivity, b) freeing up resources to be used for more productive purposes, c) reducing the need to devote scarce resources to purely maintenance functions, d) helping to coordinate activities within and across work groups, e) strengthening the organization’s ability to attract and retain the best employees, f) increasing the stability of the organization’s performance, and g) enabling the organization to more effectively adapt to environmental changes.

Consequently, there has been considerable interest in the OCB and citizenship performance literature attempting to validate these assertions. Karambeyya (1991) initiated this stream of research by examining the relationship between subjective
measures of work-unit performance and unit members’ OCB. Using a sample of 18 work units from white-collar professions, she found that high performing work units were more likely to have members who displayed OCBs than low performing units. Critical of the subjective performance criteria used in Karambayya’s study, researchers focused on replicating her findings using objective indicators of performance.

Podsakoff and MacKenzie (1997) reviewed four studies examining the relationship between OCB and objective accounts of unit level performance. These four studies considered several OCB dimensions conceptualized by Organ (1988) and investigated their relationship to quantitative and qualitative work-group performance across diverse blue-collar and white-collar jobs (MacKenzie, Podsakoff, & Ahearne, 1996; Podsakoff, Ahearne, & MacKenzie, 1997; Podsakoff & MacKenzie, 1994; Walz & Niehoff, 1996). These studies have found general support for the positive relationship between OCB and organizational effectiveness. Podsakoff and MacKenzie (1997, p. 142) provide a detailed summary of these findings:

Across four diverse samples, OCBs accounted for an average of approximately 19% of the variance in performance quantity, over 18% of the variance in the quality of performance, about 25% of the variance in financial efficiency indicators, and about 38% of the variance in customer service indicators.

The researchers also reported variance in the strength of the relationship between OCB and organizational effectiveness contingent upon the particular OCB dimension. For instance, helping behaviors demonstrated a more consistent effect (with the exception of insurance sales teams) whereas sportsmanship and civic virtue were more limited in their
Citizenship and Task Performance Ratings

relation to organizational effectiveness. Taken together, these findings provide support for Organ’s original conjecture that OCB is related to organizational performance.

Evidence for the influence of OCB on unit level performance has been well established and replicated across at least five studies. It would follow that OCB would also contribute to performance at the individual level. The OCB literature has shown support for this inference. Motowidlo and Van Scotter (1994) found that independent task and citizenship performance ratings for 300 entry-level Air Force employees correlated .43 and .41 with overall performance ratings, respectively. Their findings have suggested that citizenship performance contributes about as much as task performance to supervisory judgments of performance. Similar findings have been reported from several other studies, each supporting the argument that citizenship performance and task performance are commensurate in predicting an employee’s overall performance evaluation (Borman, White, & Dorsey, 1995; MacKenzie, Podsakoff & Fetter, 1991; Van Scotter & Motowidlo, 1996). More recently, research has explored the mediating contingencies of the relationship between citizenship performance and judgments of overall performance. Allen and Rush (1998) found that liking and perceived affective commitment mediated the relationship in two diverse samples. For both students and managers, the influence of OCB on overall ratings operated through the rater’s liking of the target subordinate or their perception of subordinate commitment.

While research has been successful in establishing a link between OCB and both subjective and objective evaluations of performance, studies have also provided support for the effects of OCB on the distribution of organizational rewards. Van Scotter,
Motowidlo, and Cross (2000) found that citizenship performance was related to promotability ratings and the attainment of informal systemic rewards for two large military samples. These relationships remained significant when experience and task performance were controlled through hierarchical regression. Allen and Rush (1998) found similar results linking OCB to recommendations for salary increase, promotion, high profile projects, public recognition, and opportunities for professional development. Accordingly, OCB has been shown to not only enhance organizational effectiveness, it also facilitates employees in the acquisition of organizational rewards and in efforts toward advancing one’s career.

**OCB, Task Performance, and Rater Accuracy**

Despite an abundance of research linking OCB to ratings of overall employee performance, the majority of formal performance appraisal systems fail to measure or consider dimensions of performance beyond the realm of in-role task requirements (Welbourne, Johnson, & Erez, 1998). Under such constraints, managers are forced to either ignore the relevance of OCB in evaluation or allow perceptions of subordinate OCB to influence task or overall performance ratings. If OCB is dismissed as a relevant performance dimension, performance ratings will fail to capture the entire performance domain and overall ratings will lack a citizenship component. However, research has shown the effect of citizenship performance on overall performance ratings (Borman, White, & Dorsey, 1995; MacKenzie, Podsakoff & Fetter, 1991; Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo, 1996). Accordingly, it is unlikely that
supervisory judgments of performance disregard OCBs, even when appraisal systems fail
to formally assess citizenship performance.

The alternative to ignoring the relevance of citizenship performance in formal
evaluations is to allow perceptions of OCB to influence performance ratings outside the
domain of citizenship, for example, in task performance ratings. While this would allow
for a more representative assessment of the entire performance domain, as a consequence,
the accuracy of dimensional (e.g. task) performance ratings would suffer. Both of these
phenomena are consistent with Wyer and Srull’s (1989) model of person memory.
According to their theory, managers store information on the dimensional and general
level concerning subordinate performance. Both sources of information are retrieved
from memory when making performance judgments. Subordinate behavior beyond a
manager’s dimensional paradigm of performance may incrementally influence the
manager’s general impressions of the subordinate. Thus, if evaluation efforts fail to
recognize citizenship performance, dimensional impressions of OCB could result in the
inflation or deflation of overall performance impressions. This reasoning, coupled with
research support for a relationship between OCB and overall performance ratings,
provided the basis for our first hypothesis (Borman, White, & Dorsey, 1995; MacKenzie,
Podsakoff & Fetter, 1991; Motowidlo & Van Scotter, 1994; Van Scotter & Motowidlo,
1996).

*Hypothesis 1(a):* The targets’ ratings of overall performance will be positively
related to the level of OCB displayed by the target.
Legally defensible performance appraisal systems must be validated through job analysis procedures (Bernardin, Kane, Ross, Spina, & Johnson, 1995). Traditional job analysis methodology focuses on in-role task behaviors while sometimes ignoring performance dimensions beyond the scope of task activities. Though researchers have realized the need to look beyond task behaviors when assessing performance, managers may pay more attention to task behaviors because they are consistent with performance appraisal standards. Although perceptions of OCB may infiltrate judgments of overall performance, raters’ perceptions of overall performance may not be influenced by citizenship behaviors if targets fail to display a minimal level of task performance. Consistent with this logic, Werner (1994) found that personal support had more of an influence on overall performance ratings when the ratee also exhibited high task performance. We expected this interaction to be replicated in the current study and apply to the entire citizenship performance domain.

**Hypothesis 1(b):** The relationship between OCB and overall performance ratings will be moderated by the targets’ level of task performance. Specifically, OCB will have a weaker effect on overall performance ratings under conditions of targets’ low task performance.

A hypothetical line graph demonstrating the interaction predicted in hypothesis 1(b) is presented in Figure 1.
Research has provided strong support for the influence of OCB on overall performance ratings; however, the OCB literature has overlooked the possible link between citizenship performance and ratings of task performance. Wyer and Srull’s theory argues that behaviors with deficient categorical outlets are grouped into available behavioral dimensions that are most representative or similar. It would follow that perceptions of OCBs, failing to fit neatly into prescribed performance dimensions, would be grouped into relevant task dimensions or overall performance judgments during the evaluation process. The OCB literature has made a strong case for distinguishing between citizenship and task performance. For instance, Conway (1996) demonstrated that a panel of industrial-organizational psychologists could reliably sort 85% of
performance dimensions across 14 studies into categories of task or citizenship performance.

While subject matter experts may be able to distinguish between citizenship and task performance, managers may have more difficulty delineating the two performance domains. Organ (1988) has argued that OCBs benefiting the organization “straddle the boundary” between citizenship and task performance. It is possible that managers often confuse behaviors promoting organizational support for in-role task requirements. Considering Organ’s argument and the postulates of Wyer and Srull, we predicted that citizenship performance will also be related to task performance ratings.

_Hypothesis 2(a):_ The targets’ ratings of task performance will be positively related to the level of OCB displayed by the target.

Recognition of citizenship within performance appraisal systems would allow managers to channel perceptions of OCB into the proper categories without sacrificing the accuracy of task performance ratings. Formal evaluation of citizenship dimensions would counteract the tendency to accommodate perceptions of OCB with incremental adjustments in ratings of task performance. Including citizenship performance dimensions in evaluation would not only provide a better representation of the performance domain, it should also increase the accuracy of task performance ratings.

_Hypothesis 2(b):_ The relationship between OCB and task performance ratings will be moderated by performance rating format. Specifically, the targets’ ratings of task performance will only be positively related to the level of OCB displayed by the target when OCB is not evaluated.
A hypothetical line graph demonstrating the interaction predicted in hypothesis 2(b) is presented in Figure 2.

![Graph showing the hypothesized interaction between citizenship performance level and rating format as a function of task performance ratings.]

*Figure 2.* The hypothesized interaction between citizenship performance level and rating format as a function of task performance ratings.

The inclusion of citizenship dimensions in performance evaluations should serve to reduce raters’ tendency to make inaccurate adjustments to task performance ratings in response to OCB perceptions. However, formal recognition of citizenship would not be expected to change impressions of overall performance. Overall performance encompasses all behaviors, task or otherwise, necessary for the successful performance of a job. Motowidlo, Borman, and Schmidt (1997, pg. 72) define overall job performance as the “aggregated value to the organization of the discrete behavioral episodes that an
individual performs over a standard interval of time.” They argue that as much as 30% of managerial performance is accounted for by citizenship behaviors. If citizenship is a requisite of successful performance, OCB should influence overall performance ratings inadvertent of the dimensional considerations of performance appraisals. This logic was the basis for our final hypothesis.

_Hypothesis 3:_ The targets’ ratings of overall performance will not differ as a function of rating format.
Method

Performance Constructs and Measures

Citizenship Performance. The current study utilized mock supervisory logs to manipulate levels of OCB consistent with Coleman and Borman’s (2000) three-dimension model of citizenship performance. Dimensions of organizational support, personal support, and conscientious initiative were manipulated to create three performance levels for each facet (low, moderate, and high performance). Behavioral definitions for each of the citizenship dimensions are provided in Table 1.

Citizenship performance was measured using Motowidlo and Van Scotter’s (1994) 15-item scale of OCB (See Appendix A). Raters judged the likelihood that targets would engage in acts of citizenship performance using a five-point Likert scale (1 = not at all likely, 5 = extremely likely). Motowidlo and Van Scotter reported a Cronbach reliability estimate of .95; in this study, the estimate was .94. Although this measure is based on the five-factor model of contextual performance, its items are consistent with Coleman and Borman’s (2000) three-factor taxonomy of citizenship performance. Average item scores for the 15-item scale were treated as the unit of analysis.
Table 1

*Coleman and Borman’s (2000) Taxonomy of Citizenship Performance*

<table>
<thead>
<tr>
<th>A. Personal Support</th>
<th>B. Organizational Support</th>
<th>C. Conscientious Initiative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helping others by offering suggestions, teaching them useful knowledge or skills, directly performing some of their tasks, and providing emotional support for their personal problems. Cooperating with others by accepting suggestions, informing them of events they should know about, and putting team objectives ahead of personal interests. Showing consideration, courtesy, and tact in relations with others as well as motivating and showing confidence in them.</td>
<td>Representing the organization favorably by defending and promoting it, as well as expressing satisfaction and showing loyalty by staying with the organization despite temporary hardships. Supporting the organization’s mission and objectives, complying with organizational rules and procedures, and suggesting improvements.</td>
<td>Persisting with extra effort despite difficult conditions. Taking the initiative to do all that is necessary to accomplish objectives even if not normally a part of own duties, and finding additional productive work to perform when own duties are completed. Developing own knowledge and skills by taking advantage of opportunities within the organization and outside the organization using own time and resources.</td>
</tr>
</tbody>
</table>
| Subdimensions: Helping  
Cooperating  
Courtesy | Subdimensions: Representing  
Loyalty  
Compliance | Subdimensions: Persistence  
Initiative  
Self-Development |
Task Performance. Supervisory logs also manipulated performance levels across task performance dimensions. Three dimensions of task performance were used to balance supervisory logs with an equal number of OCB and task behavior statements. Task performance dimensions included Task Proficiency/Quality, Production/Efficiency, and Judgment/Problem Solving. These task performance dimensions were manipulated to create three performance levels for each facet (low, moderate, and high performance). Construct definitions of high performance in each of the task dimensions are listed below. Definitions were written using Borman, Ackerman, and Kubisiak’s descriptions of general task performance dimensions (1994).

Task Proficiency/Quality – Displaying a mastery of work tasks; demonstrating accuracy in own work; giving attention to detail and avoiding making mistakes and errors; producing a high quality standard of work.

Productivity -- Using resources effectively and operating in a cost-effective manner; using time well and meeting deadlines under any circumstance; consistently producing large amounts of work.

Judgment and Problem Solving – Making good decisions when faced with problems or obstacles; accurately analyzing situations or problems and determining the correct course of action given the information available; successfully solving problems by making informed decisions.

It should be noted; the three task performance dimensions chosen for this study are not exhaustive of the task performance domain. However, they are consistent with the important functions of the job simulated in this study (i.e. Financial Manager). O’Net
(2002) recognizes elements of task proficiency, production, and problem solving among the most important skill requirements under the job title “Financial Managers, Branch or Department.”

Task performance was measured with 9 items developed using behavioral task statements demonstrating high task performance (See Appendix A, items 2, 3, 4, 5, 6, 8, 9, 11, & 12). Three items were written for each of the task performance dimensions being assessed in this study. Raters judged the likelihood that targets would engage in acts of task performance using a five-point Likert scale (1 = not at all likely, 5 = extremely likely). The reliability estimate for the scale was .93. Average item scores for the 9 task items were used as the unit for analysis.

**Overall Performance.** Three items were developed to assess the overall evaluation of the ratee. Using a five-point Likert scale (1 = not at all likely, 5 = extremely likely), raters judged the likelihood that targets’ future performance would demonstrate overall excellence (See Appendix B, items 1, 7, & 10). One overall performance item (Item 7) was reverse scored. The reliability estimate for the scale was .76. Average item scores across the three overall performance items were used as the unit for analysis.

**Sample**

Participants consisted of 360 undergraduate psychology students enrolled in psychology courses at a large southeastern university. The sample consisted of 78 males and 282 females. Participants’ ages ranged between 18 and 50 years with a median of 21 years. All participants were compensated with extra credit in psychology courses.
Likewise, all participants read and signed an informed consent before participating in the experiment.

**Procedure**

All experimental manipulations were performed on paper via supervisory logs and rating scales. Random assignment to experimental conditions was achieved by randomly ordering material sets before distribution and administration. This allowed for data to be collected from groups of participants in a classroom setting while preserving random assignment. All participants received the same introduction to the experiment and the same instructions pertaining to the rating scales. Participants were not trained on rater error, accuracy, or frame-of-reference.

Each participant reviewed a single supervisory log and rated the target’s performance immediately after reading the log. As prescribed by the between subjects factorial model, each participant was exposed to a single level of task performance, a single level of citizenship performance, and one rating format. Upon completion of the rating scale, all materials were collected and participants were debriefed simultaneously. Data collection sessions did not exceed 15 minutes.

**Supervisory Logs**

Supervisory logs were developed using work-related behaviors of first-line financial managers (See Appendix C). Logs were written to simulate the documentation of observed subordinate behavior by a divisional director of a financial management firm. Each log contained 12 statements describing a target’s work quality in each of the performance dimensions (6 citizenship performance statements, 6 task performance statements).
statements). Citizenship performance and task performance subdimensions were each represented with two behavioral statements per supervisory log. All 6 behavioral statements of citizenship performance corresponded to a single level of performance within each log. Likewise, all 6 task performance statements corresponded to a single level of performance in each of the 9 logs. The name and gender of the subordinate was kept constant across all 9 logs.

Behavioral statements were adapted from Borman, Ackerman, and Kubisiak’s technical report (1994) to reflect three levels of performance across both task and citizenship performance dimensions. A total of 119 behavioral statements of citizenship performance and 82 statements of task performance were developed. Four industrial/organizational psychology graduate students trained in performance assessment served as expert raters. These raters categorized both citizenship and task statements according to dimension. They also rated each behavioral statement in terms of its relative level of effectiveness. Expert raters used a five-point Likert scale to assess each statement (5 = Exceptional performance, 4 = Good performance, 3 = Average performance, 2 = Below average performance, 1 = Poor performance). Behavioral statements failing to achieve 75% accuracy on dimension sorting across expert raters were disqualified from use in supervisory logs. Furthermore, mean expert ratings for each statement greater than .50 scale points from the intended performance level were also disqualified (Low = 1 - 1.5, Medium = 2.5 – 3.5, High = 4.5 – 5). A total of 14 behavioral statements failed to meet one or both of the established standards of inter-rater agreement.
Supervisory logs were developed with the remaining 187 statements meeting the qualification criteria. After each log had been assigned the appropriate number of statements, situational stems were added to the behavioral statements to enhance the job simulation (e.g. At the board meeting John “insert behavioral statement”). These stems were randomly assigned to statements within logs and remained constant across all 9 logs. After development of the supervisory logs, the four expert raters sorted the logs into nine categories representing all possible combinations of task and citizenship performance. Raters reached perfect consensus on the appropriated categorization of the 9 supervisory logs on their first attempt. Accordingly, log revisions were not necessary.

Design

The current study utilized a 3x3x2 between subjects full factorial model. The three independent variables included citizenship performance (three levels), task performance (three levels), and rating format (inclusion of OCB, or exclusion of OCB). Participants rated the target’s overall and task performance, or they rated the target’s overall, task, and citizenship performance. A total of 20 participants were randomly assigned to each of the 18 cells of the factorial model.
Results

Two separate multivariate analyses of variance (MANOVA) were performed to test all five study hypotheses. The probability of Type I error was maintained at .05 for significance tests of the F statistic. Fisher’s LSDs were computed for all post hoc mean comparisons.

*Overall Performance Ratings*

The first MANOVA treated overall performance ratings as the dependent variable and tested the effects of task performance level, citizenship performance level, the interaction of task and citizenship performance level, and rating format (See Table 2 for summary statistics). There were significant main effects for both task performance level and citizenship performance level, $F(2, 350) = 134.90$, $MSe = 0.56$; $F(2, 350) = 36.28$, $MSe = 0.56$, respectively. As can be seen in Figure 3, overall performance ratings increased as the targets’ level of task performance and citizenship performance were increased. A subsequent Fisher’s LSD of 0.19 indicated that this positive relationship was stable across all levels of both citizenship and task performance. The positive relationship between citizenship performance and overall performance ratings provides support for hypothesis 1(a).
Hypothesis 1(b) predicted that this relationship would be moderated by task performance level. Specifically, we predicted that citizenship performance would have less of an effect on overall performance ratings under conditions of low task performance. However, there was no significant effect for the interaction term of citizenship and task performance level, $F(4, 350) = 0.36, MSe = 0.56$. The effect of citizenship performance level on overall performance ratings was constant across all levels of task performance. A lack of an interaction fails to support hypothesis 1(b).

Hypothesis 3 made a null prediction that the inclusion of OCBs in the performance assessment tool would not affect ratings of overall performance. Accordingly, we assumed that rating format would not influence participants’ perceptions.
of targets’ overall performance. In support of this assertion, our findings indicate that rating format did not affect overall performance judgments, \( F(1, 350) = 3.10, \text{MSe} = 0.56 \). It should be noted, however, because this prediction was conveyed in terms of a null hypothesis our results do not statistically disconfirm a relationship between rating format and overall performance ratings.

Table 2

Summary Table for Overall Performance Ratings MANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>194.16</td>
<td>9</td>
<td>21.57</td>
<td>38.55*</td>
<td>.0001</td>
</tr>
<tr>
<td>Citizenship Level</td>
<td>40.61</td>
<td>2</td>
<td>20.31</td>
<td>36.28*</td>
<td>.0001</td>
</tr>
<tr>
<td>Task Level</td>
<td>151.00</td>
<td>2</td>
<td>75.50</td>
<td>134.90*</td>
<td>.0001</td>
</tr>
<tr>
<td>Citizenship*Task Task</td>
<td>.81</td>
<td>4</td>
<td>.20</td>
<td>.36</td>
<td>.84</td>
</tr>
<tr>
<td>Rating Format</td>
<td>1.74</td>
<td>1</td>
<td>1.74</td>
<td>3.10</td>
<td>.08</td>
</tr>
<tr>
<td>Error</td>
<td>195.88</td>
<td>350</td>
<td>.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>390.04</td>
<td>359</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* \( p < .01 \)

Task Performance Ratings

The second MANOVA treated task performance ratings as the dependent variable and tested the effects of citizenship performance level, rating format, and the interaction term for citizenship performance and rating format (See Table 3 for summary statistics). There were significant main effects for citizenship performance level, \( F(2, 354) = 9.26, \text{MSe} = 0.94 \). Consequently, there was a positive relationship between citizenship performance level and ratings of task performance. This finding provides support for hypothesis 2(a).
Hypothesis 2(b) predicted that the effects of citizenship level on task performance ratings would be moderated by rating format. Specifically, it was predicted that the effect of citizenship performance on task performance ratings would be reduced when citizenship performance was included in the assessment tool. However, there was no significant effect for the interaction of citizenship performance and rating format on ratings of task performance, $F(2, 354) = 0.02$, $MSe = 0.94$. As can be seen in Figure 4, the positive relationship between citizenship performance and task performance ratings is evident across both rating formats. Consequently, hypothesis 2(b) was not supported.
Table 3

Summary Table for Task Performance Ratings MANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>18.58</td>
<td>5</td>
<td>3.72</td>
<td>3.97*</td>
<td>.002</td>
</tr>
<tr>
<td>Citizenship Level</td>
<td>17.32</td>
<td>2</td>
<td>8.66</td>
<td>9.26*</td>
<td>.0001</td>
</tr>
<tr>
<td>Rating Format</td>
<td>1.23</td>
<td>1</td>
<td>1.23</td>
<td>1.31</td>
<td>.25</td>
</tr>
<tr>
<td>Citizenship*Format</td>
<td>.04</td>
<td>2</td>
<td>.02</td>
<td>.02</td>
<td>.98</td>
</tr>
<tr>
<td>Error</td>
<td>331.09</td>
<td>354</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>349.68</td>
<td>359</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .01

Post hoc analyses were conducted to determine if the positive relationship between citizenship performance and task performance ratings was stable across all citizenship performance levels. After pooling scores across rating formats, F(2, 359) = 9.30, MSe = .93, p<.0001, a subsequent Fisher’s LSD of 0.25 indicated that citizenship performance only influenced task performance ratings under conditions of high citizenship performance (See Figure 5).
Figure 5. The effects of citizenship performance level on ratings of task performance pooled across rating formats.
Discussion

Implications

The results provided partial support for the study’s hypotheses. Both citizenship and task performance levels positively influenced judgments of overall performance. This finding is consistent with prior findings (Borman, White, & Dorsey, 1995; MacKenzie, Podsakoff & Fetter, 1991; Van Scotter & Motowidlo, 1996). Interestingly, though, Werner’s (1994) finding that this relationship was moderated by task performance level was not replicated. Our results indicate that perceptions of citizenship performance predict overall performance equally well across all task performance levels.

It should be noted, however, that Werner’s study only considered the interactive effects of task performance and personal support on performance judgments. The current study manipulated citizenship across its entire performance domain capturing facets of conscientious initiative, personal support, and organizational support. It may be the case that raters were unable to consistently distinguish between citizenship and task performance behaviors when evaluating OCBs beyond the dimension of personal support. As mentioned earlier, Organ (1988) argued that OCBs, reflecting organizational support, are more likely to “straddle the boundary” of in-role and extra-role behaviors. Similar arguments have been made by other researchers studying the OCB performance domain (VanDyne & Cummings, 1990). This boundary may become increasingly muddied for managerial jobs involving complex responsibilities.
Although OCB theorists now acknowledge that OCBs may be in-role or extra-role behaviors (Organ, 1997), those evaluating the entire citizenship performance domain may consider such behaviors material to overall performance ratings regardless of an employee’s level of task performance. Consistent with this assertion, our findings indicate that there is no minimally acceptable task performance requisite before citizenship behaviors are incorporated in overall performance judgments. This provides support for the argument that citizenship and task performance are commensurate in determining overall performance ratings.

Another major finding of this study was the positive relationship between citizenship performance level and task performance ratings. Post hoc analyses indicated that task performance ratings were inflated only when citizenship performance was high. This result is consistent with Werner’s (1994) finding that raters are significantly more likely to commit halo error when targets display high levels of OCB than when targets display neutral levels of OCB. Surprisingly, however, the inclusion of OCBs in the rating scale did not increase the accuracy of the task performance ratings. In fact, though not statistically significant, task performance ratings were slightly overestimated when citizenship performance was assessed.

Following from Wyer and Srull’s (1989) person memory theory, we predicted that raters would not incorporate perceptions of citizenship performance in task performance ratings when provided the opportunity to evaluate citizenship in the performance assessment. As such, raters would be provided an outlet to assess OCBs within relevant performance categories without sacrificing the accuracy of task performance ratings.
One explanation for why our results failed to support this reasoning may be that raters were able to distinguish between task and citizenship performance behaviors regardless of whether citizenship was included in the assessment scale. According to this explanation, the person memory theory would not apply to performance evaluation.

An alternative explanation is that instead of adjusting task performance ratings in lieu of ratees’ citizenship performance levels, raters incorporated perceptions of targets’ citizenship into overall performance judgments. As mentioned earlier, the person memory theory contends that performance behaviors with deficient categorical outlets are grouped into the most representative or similar behavioral dimensions being assessed. Conway’s (1996) research indicates that experienced raters can successfully distinguish between task and citizenship performance dimensions. Thus, the distinction between task and citizenship performance dimensions may deter raters from considering perceptions of targets’ citizenship when making task performance judgments. However, the same logic would not apply for the assessment of overall performance. Because the overall performance domain collectively represents all facets of performance, the influence of citizenship perceptions on overall performance judgments is not only warranted but expected. In fact, perceptions of citizenship performance should influence raters’ judgments of overall performance whether or not citizenship is included in the appraisal scale. This is evident by our finding that overall performance ratings were unaffected by rating format. While the inclusion of OCBs in performance rating scales may add richness to performance ratings for job placement or developmental purposes, the current
findings suggest that the accuracy of task and overall performance ratings is not influenced.

Limitations

One limitation of this study was the utilization of a fixed-effects design. This research carefully constructed two primary performance dimensions (citizenship & task) and manipulated three performance levels while maintaining performance continuity across performance subdimensions (e.g. personal support, organizational support, conscientious initiative). The likelihood that a single level of citizenship performance or task performance behaviors would be displayed by an employee in a real-world setting is questionable. Following the procedural assumptions of Kirk (1982), study conclusions only apply to the treatment levels used in this experiment. It is possible that results could vary given other combinations of performance levels within the task or citizenship domain.

A second limitation of this study was the use of hypothetical “paper” employees as targets and undergraduate psychology students as raters. The assessment task used in this experiment was relatively simple. Participants were required to make quick judgments of targets’ performance based on only twelve behavioral examples. Although the use of vignettes and mock supervisory logs allow for experimental manipulation of performance levels, they may not capture some of the contextual cues involved in making performance judgments in an organizational work setting. Similarly, an argument could be made that undergraduate students, serving as raters, are not representative of real-
world supervisors evaluating employee performance. Differences may exist between student and supervisory conceptualizations of the performance domain.

**Conclusion**

As has been evidenced in previous studies, this research found that both citizenship and task behaviors are important in determining overall performance judgments. However, this study also demonstrated that levels of citizenship performance have an effect on ratings of task performance. Finally, our results indicate that including citizenship dimensions in the appraisal format does not affect the accuracy of task or overall performance ratings. However, this isn’t to say that performance assessment should dismiss the potential advantages of evaluating citizenship performance. By including OCBs in formal evaluations, employers can gain more complete information of employee performance across a wider behavioral range. Richer performance appraisal information can benefit employers in the areas of employee placement, retention, and development.
Reference List


Appendix A. Motowidlo and Van Scotter’s (1994) 16 item scale of Citizenship Performance

Please rate the manager by responding to each statement with the most appropriate answer:

1 = not at all likely  2 = not likely  3 = somewhat likely  4 = likely  5 = extremely likely

**While performing his or her job, how likely is it that this person would…**

<table>
<thead>
<tr>
<th></th>
<th>Not at Likely</th>
<th>Extremely Likely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Comply with instructions even when supervisors are not present.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>2. Cooperate with others in the team.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>3. Persist in overcoming obstacles to complete a task.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>4. Display proper company appearance and manner.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>5. Volunteer for additional responsibilities.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>6. Follow standard operating procedures and avoid unauthorized shortcuts.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>7. Look for challenging assignments.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>8. Offer to help others accomplish their work.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>9. Pay close attention to important details.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>10. Defend the supervisor’s decisions.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>11. Render proper business courtesy.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>12. Support and encourage a coworker with a problem.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>13. Take the initiative to solve a work task.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>14. Exercise personal discipline and self-control.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>15. Tackle a difficult work assignment enthusiastically.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>16. Voluntarily do more than the job requires to help others or contribute to company effectiveness.</td>
<td>1 2 3 4 5</td>
<td>1 2 3 4 5</td>
</tr>
</tbody>
</table>
Appendix B. Task and Overall Performance Scale

Please rate the manager by responding to each statement with the most appropriate answer:

\[ 1 = \textit{not at all likely} \quad 2 = \textit{not likely} \quad 3 = \textit{somewhat likely} \quad 4 = \textit{likely} \quad 5 = \textit{extremely likely} \]

While performing his or her job, how likely is it that this person would…

<table>
<thead>
<tr>
<th>Statement</th>
<th>Not at Likely</th>
<th>Extremely Likely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Perform at a level much higher than coworkers.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>2. Produce a large amount of work.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>3. Accurately analyze situations and determine the correct course of action.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>4. Display a mastery of work tasks.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>5. Make informed decisions.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>6. Pay attention to detail and avoid making mistakes.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>7. Consistently perform below work standards.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>8. Use resources in a cost-effective manner.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>9. Make good decisions in the presence of obstacles.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>10. Act as the best employee under your supervision.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>11. Produce a high quality standard of work.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>12. Meet deadlines under any circumstance.</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C. Annual Supervisory Report Log

Employee: __________________________  Supervisor: __________________________
Department: ______________________  Job Title: _____________________

<table>
<thead>
<tr>
<th>Observation</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>At the department staff meeting Bill offered sound suggestions for changes in administrative and organizational procedures that would better serve the company’s mission and objectives. (OS)</td>
<td>Jan. 14</td>
</tr>
<tr>
<td>When preparing for the last quarterly report Bill could be counted on to take additional tasks when others asked for help due to being overloaded. (PS)</td>
<td>Mar. 3</td>
</tr>
<tr>
<td>On his last group assignment Bill voluntarily performed tasks that are not normally a part of his duties when necessary. (C)</td>
<td>Mar. 29</td>
</tr>
<tr>
<td>It seems that Bill demonstrates knowledge of his position, skills needed to accomplish tasks, and the ability to perform those skills. (Q)</td>
<td>May 2</td>
</tr>
<tr>
<td>I’ve noticed that Bill independently thinks through problems and creatively pursues a resolution while allowing for more than one solution. (J)</td>
<td>June 11</td>
</tr>
<tr>
<td>Bill accomplishes job assignments/tasks quickly while using the minimum amount of resources possible. (P)</td>
<td>July 19</td>
</tr>
<tr>
<td>It is apparent that Bill takes the initiative to correct obviously non-standard conditions. (C)</td>
<td>Aug. 8</td>
</tr>
<tr>
<td>Bill’s coworkers have told me that he gives encouragement when approached by workers who are experiencing adversity or setbacks. (PS)</td>
<td>Aug. 31</td>
</tr>
<tr>
<td>Bill tends to effortlessly exceed requirements for amount of work performed. (P)</td>
<td>Oct. 14</td>
</tr>
<tr>
<td>It is evident in Bill’s reports that he understands a problem/situation and takes the necessary steps to correct the problem. (J)</td>
<td>Nov. 4</td>
</tr>
<tr>
<td>It is clear that Bill actively embraces the organization's missions and objectives. (OS)</td>
<td>Nov. 19</td>
</tr>
<tr>
<td>Bill performs tasks to achieve quality goals/standards, thoroughly understanding the need for quality. (Q)</td>
<td>Dec. 9</td>
</tr>
</tbody>
</table>

Supervisor Signature: __________________________  Date: __________________________

Researcher notes:
C = Conscientiousness  J = Judgment/Problem Solving
OS = Organizational Support  P = Productivity
PS = Personal Support  Q = Quality/Task Proficiency

*This is an example of a log for high-citizenship/high-task performance.
**Bold words indicate sentence stems for behavioral statements