Managerial ratings of in-role behaviors, organizational citizenship behaviors, and overall performance: testing different models of their relationship

Kevin Barksdalea,1, Jon M. Wernerb,*

aVice-President, Chuck Hutton Chevrolet Co., Memphis, TN 38115, USA
bDepartment of Management, College of Business and Economics, University of Wisconsin-Whitewater, 800 West Main Street, Whitewater, WI 53190-1790, USA

Accepted 1 March 2000

Abstract

Recent writings by Borman and Motowidlo [Borman WC, Motowidlo SJ. Expanding the criterion domain to include elements of contextual performance. In: Schmitt N, Borman WC, editors. Personnel Selection in Organizations. San Francisco: Jossey-Bass, 1993. pp. 71–98], Organ [Organ DW, Organizational citizenship behavior: it’s construct clean-up time. Hum Perform 1997; 10: 85–97], and others have sought to clarify or reposition organizational citizenship behavior (OCB) as an element of a broadened definition of employee performance. We utilized confirmatory factor analysis to test several of the assumptions behind this new direction for OCB research. First, in support of prior research, we found that in-role behaviors (IRBs) and two dimensions of OCB (altruism and conscientiousness) were empirically distinct. Next, we found that overall performance ratings were predicted by ratings given concerning IRB and altruism, though not by ratings of the OCB dimension of conscientiousness. Third, a second-order factor analysis that specified four first-order factors loading on one general factor of performance was found to be consistent with the data. This is presented as support for including OCB dimensions within current definitions of employee performance. Finally, to address possible halo in the data, a second causal model was evaluated, where overall performance was viewed as causally prior to the other three measures. Implications are discussed. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: In-role behavior; Organizational citizenship behavior; Overall performance; Halo

The past decade has seen a large amount of research and conceptual development concerning organizational citizenship behavior, or OCB (Smith et al., 1983; Organ, 1988, 1990, 1994, 1997; Organ and Konovsky, 1989; MacKenzie et al., 1991; Moorman, 1991; Van Dyne et al., 1994). Even though OCB was originally presented as “a category of performance” (Smith et al., 1983, p. 653), much of the subsequent research on this topic has proceeded without consideration as to how OCB fits within the larger domain of performance. Recently, however, Waldman (1994) used a total quality management framework to define work performance. Waldman specifically included citizenship behaviors within his definition of work performance, since such behaviors as cooperation and taking initiative “above and beyond the call of duty” are likely to be associated with unit-level measures of performance, especially in organizations with strong quality efforts in place.

In a different vein, Campbell (1990) and Campbell et al. (1993) presented a performance taxonomy, where performance was viewed as multi-dimensional. Interestingly, of the eight factors proposed by Campbell and his associates, three factors overlap highly with dimensions of organizational citizenship, i.e., demonstrating effort, facilitating peer and team performance, and maintaining personal discipline. Along these same lines, some researchers have argued that citizenship dimensions should be viewed as sub-factors within a general performance construct (Borman and Motowidlo, 1993; Motowidlo and Van Scotter, 1994; Werner, 1994, in press; Murphy and Shiarella, 1997). Other researchers, however, have argued that performance is best viewed in terms of one general factor (Viswesvaran et al., 1993), something akin to “g” in the research on intelligence
and cognitive ability. Obviously, much more empirical work is needed on the performance construct; Campbell (1993), for example, has described our current state of knowledge in this area as “primitive.” The purpose of our study was to both replicate and extend previous research that has compared OCB with other aspects of employee performance. Previously proposed theoretical models of these interrelationships were tested using structural equations modeling. The following questions were addressed by our research: Can managers distinguish between in-role behavior (IRB) and OCB? Do managers use information on both types of behaviors when forming overall performance ratings? Can ratings of various aspects of employee performance be seen as falling under a general or higher-order performance factor, or alternately, are the intercorrelations more likely to be due to a halo effect, i.e., general impressions in the minds of raters that color all ratings given?

Katz and Kahn (1978) distinguished between job-related behaviors that they labeled in-role and extra-role. IRB has been defined as behavior that is required or expected as part of performing the duties and responsibilities of an assigned work role, whereas extra-role behavior is discretionary behavior that benefits the organization and that goes beyond existing role expectations (Van Dyne et al., 1995). These categories have also been labeled as “core” and “discretionary” behaviors (Tompson and Werner, 1997), and a similar but not identical distinction has been made between task and contextual performance (Borman and Motowidlo, 1993). Contextual performance is described as a “set of interpersonal and volitional behaviors that support the social and motivational context in which organizational work is accomplished” (Van Scotter and Motowidlo, 1996, p. 525). Whereas a good deal of work has been done concerning dollar-based evaluations of employee performance (Cascio, 2000), and in particular, how individual performance contributes to organizational success (Ostroff, 1992), holistic evaluations of employee behaviors continue to play an important role in organizational evaluations of employees.

In a recent reappraisal of the OCB construct, Organ (1997) proposed that OCB be equated with Borman and Motowidlo’s (1993) concept of contextual performance. Organ then called for more research demonstrating both convergent and discriminant validity between OCB and other performance measures. Our study addresses some of these measurement issues. We present no new measure of performance (either in-role or OCB). However, if in this study OCB is shown to be both distinguishable from IRB, and important in its own right in determining overall evaluations of employee performance, then an implication of our research is that most current performance measures are deficient in not explicitly including measures of organizational citizenship (or related constructs; Van Scotter and Motowidlo, 1996). In order to frame organizational citizenship within this broader performance construct, we next briefly describe and review two classes of job behaviors, i.e., core (in-role) and OCBs.

1. Core task (in-role) behaviors

Core or IRB were initially described by Katz and Kahn (1978) as those behaviors that are prescribed and defined as being part of one’s job, and are recognized by the organization’s formal reward systems. Williams and Anderson (1991) defined IRB by behaviors such as working a full 8-h day, or completing all required assignments. Performance criteria to measure in-role performance are often broken down into four basic categories: ratings, quality measures, quantity measures, and file data, such as records concerning safety, absences, or tardiness (Ghiselli and Brown, 1955). Supervisory ratings are often thought to be the least effective in capturing true job performance, as they may measure things that are not a part of true job performance (i.e., measurement contamination), or fail to measure other things that are a part of true performance (i.e., measurement deficiency; Wexley and Yukl, 1984). However, supervisory ratings remain the most popular form of performance measurement, perhaps due to the fact that performance evaluations remain critical to managerial decision-making and rewards processes (Murphy and Cleveland, 1995). Although recent evidence would suggest that supervisory ratings are influenced by both in-role and extra-role behaviors (MacKenzie et al., 1991; Werner, 1994), such ratings are often thought to be primarily capturing in-role work performance. Thus, while ratings of IRBs are generally thought to be highly related to supervisory ratings, it is not likely that the two are synonymous.

2. Organizational citizenship behaviors

While the distinction between in-role and extra-role behaviors has been traced to Barnard (1938) and Katz and Kahn (1952, 1978), Organ and his colleagues have made the strongest contribution to our understanding of discretionary or OCBs (Bateman and Organ, 1983; Smith et al., 1983; Organ, 1988, 1990; Organ and Konovsky, 1989). Organ (1990, p. 46), defined OCBs as “those organizationally beneficial behaviors and gestures that can neither be enforced on the basis of formal role obligations nor elicited by contractual guarantee of recompense.” Smith et al. (1983) developed a 16-item measure of OCB consisting of two dimensions: altruism and general compliance (also referred to as conscientiousness; Organ, 1988). Altruism is described as those behaviors that have to do with helping a specific person, (supervisor, co-worker, or client). Conscientiousness refers to behavior that goes beyond minimum standards (i.e., in attendance, use of work time, adherence to rules). The Smith et al. (1983) measure has subsequently been shown to have a fairly consistent factor structure (Smith et al., 1983; Organ and Konovsky, 1989; Shore and Wayne, 1993; Shore et al., 1995), although items have occasionally failed to load unambiguously on their

Concerns continue to be raised regarding the construct validity of OCB (Van Dyne and Cummings, 1990; Becker and Vance, 1993; Morrison, 1994; Organ, 1997). A number of different conceptualizations and measurement scales have been proposed (Organ, 1988; Graham, 1989; Podsakoff et al., 1990; Van Dyne et al., 1994; Moorman and Blakely, 1995). A particular concern is the extent to which raters of performance can distinguish OCB from other aspects of employee behavior (McAllister, 1991). For example, it has been argued that OCB, as presently measured, is often indistinguishable from organizationally prescribed roles (Morrison, 1994). However, several recent studies have demonstrated the empirical distinctiveness of OCB from other forms of employee behavior (O’Reilly and Chatman, 1986; Williams and Anderson, 1991; McNeely and Meglino, 1994).

For example, Williams and Anderson (1991) examined the relationship of employees’ IRBs and two broad categories of OCBs with organizational commitment and job satisfaction. They used the OCB factors from Smith et al. (1983), and expanded and re-labeled these as OCBI (citizen-citizenship behaviors that benefit specific individuals), and OCBO (citizenship behaviors that benefit the organization in general). Even though these factors were highly intercorrelated ($r$ values ranged from 0.52 to 0.56), these authors demonstrated via factor analysis that raters could distinguish between IRB and these two types of citizenship behaviors, i.e., measures of IRB and OCBI represented distinct, yet related factors of performance. In this study, we utilized a different sample and somewhat different measures than Williams and Anderson (1991). However, our first hypothesis is primarily intended as a replication of the findings from these earlier studies. It is important to establish the distinctiveness of these various forms of behavior before the other hypotheses in our study are addressed.

**H1. Managers will perceive IRBs, altruism, and conscientiousness as distinct, yet related aspects of employee performance.**

Other research on OCB has sought to use ratings of in-role and extra-role behaviors predictively. Here, the general question is whether managers and other raters of performance consider both aspects of employee performance when forming overall ratings of employee performance (Motowidlo and Van Scotter, 1994; Werner, 1994; Van Scotter and Motowidlo, 1996). For example, MacKenzie et al. (1991) found that supervisory ratings of insurance agent performance were significantly predicted by an objective performance measure (sales), as well as by two OCB dimensions (altruism and civic virtue). Podsakoff and MacKenzie (1994, Study 1) found that altruism, civic virtue, and sportsmanship were each significantly related to a three-item measure of overall job performance. The OCB dimension of conscientiousness was not measured in either of these studies, since behaviors captured by this dimension were “not viewed as a discretionary form of behavior in this particular context” (Podsakoff and MacKenzie, 1994, p. 354).

Werner (1994) manipulated the level of secretarial performance on both in-role and OCBI-type dimensions (i.e., altruism and extra effort). Both IRB and OCBI ratings were significantly related to the overall ratings assigned. Items tapping OCBO dimensions (such as conscientiousness) were included in Werner’s study, but were not independently manipulated, and thus the impact of conscientiousness ratings on overall performance ratings was not measured. Finally, Van Scotter and Motowidlo (1996) studied the ratings of Air Force mechanics on three variables that are conceptually very similar to IRB, altruism, and conscientiousness. They obtained the following correlations with overall performance ratings: task performance ($r = 0.56$), interpersonal facilitation ($r = 0.44$), and job dedication ($r = 0.54$).

A second purpose of the present study was to replicate and extend this line of research concerning the extent that overall performance ratings can be predicted based on managerial ratings of IRB, altruism, and conscientiousness. The above-mentioned research (MacKenzie et al, 1991; Podsakoff and MacKenzie, 1994; Werner, 1994) provides empirical support for the links between IRB and overall ratings, and altruism and overall ratings. The link between conscientiousness and overall ratings can be drawn from the theoretical work of Organ (1988).

To avoid confusion, the conscientiousness dimension of organizational citizenship should be distinguished from the personality construct by the same name. In the personality literature, conscientiousness is one of the “Big Five” dimensions that have increasingly been used to describe major categories of personality (Digman, 1990; Goldberg, 1993). As a personality construct, conscientiousness refers to characteristics such as being organized, dependable, willing to achieve, and persevering (Costa and McCrae, 1985). The potential for confusion in the contrasting of personality and behavior has not gone unnoticed (Organ, 1997). In fact, the original label for OCB-conscientiousness was generalized compliance.

Two meta-analyses shed light on the relationship of conscientiousness as a personality construct to measures of performance. First, Barrick and Mount (1991) found that of the Big Five personality dimensions, only conscientiousness had consistent and statistically significant relationships with performance across various occupational groups and criterion measures (with corrected correlations ranging from $r = 0.20$ to $0.23$). Second, Organ and Ryan (1995) found that conscientiousness correlated $r = 0.21$ ($r = 0.30$, corrected for unreliability) with the OCB dimension of conscientiousness. Thus, while there is clearly an overlap between the two constructs, they are not synonymous. The OCB measure of conscientiousness places particular emphasis on punctuality, attendance, and overall dependability, whereas the person-
Hypothesis 1 concerns the distinctiveness of IRB, altruism, and conscientiousness, and Hypothesis 2 utilizes these three variables to predict overall ratings of performance. A question of perhaps even greater interest and importance has to do with the underlying causes of the interrelationships observed between these four variables. One way of viewing these interrelationships could be labeled the “p view.” Similar to the Spearmanian notion of “g” representing a general factor of cognitive ability, this approach would suggest that at least a good portion of the observed relationship between different measures of performance is due to “p,” or a general factor of performance. Williams and Anderson (1991), for instance, proposed that future research should address the extent that in-role and citizenship dimensions might represent indicators of a general factor of performance. However, because their study only included three measures of performance, their data did not allow them to adequately test this hypothesis. If one assumes that our measure of overall performance is not simply a summation of IRB, altruism, and conscientiousness, but in fact, captures some dimensions of performance that are distinct from either of the other three scales (e.g., decision-making or problem-solving), then these four measures can be used to test a second-order model (depicted as Model 2 in Fig. 1). Given the high degree of interrelationship observed between these variables in previous research (Williams and Anderson, 1991; Shore and Wayne, 1993), the third purpose of the present research was to test the general factor of performance hypothesis presented by Williams and Anderson (1991). Therefore, our final hypothesis is as follows:

**H3. A second-order common factor will underlie IRBs, altruism, conscientiousness, and overall performance ratings.**

Although Hypothesis 3 is our preferred view, another position that could be taken is what could be called the “halo view.” Proponents of this view would argue that these variables are highly interrelated because raters form general impressions that then serve to “color” their ratings of specific performance dimensions. This topic has been studied extensively in the performance appraisal literature concerning the halo effect (Cooper, 1981; Murphy and Balzer, 1989; Balzer and Sulsky, 1992; Lance et al., 1994b).

Several techniques have been introduced for evaluating the effects of halo. For example, Landy et al. (1980) recommended statistical partialling to control for the impact of a “general impression” or general factor of performance. Similarly, Nathan and Lord (1983) utilized this approach to partial out overall ratings of performance from the intercorrelations among specific dimensions. They found that general impression halo explained a fair amount of the relationship between these dimension ratings, but that meaningful intercorrelations remained, even after a general factor had been partialled out.

However, the use of statistical partialling has been criticized by several authors (Harvey, 1982; Hulin, 1982; Murphy, 1982; Podsakoff and Organ, 1986). Podsakoff and Organ (1986) are critical of partialling as a means of determining common method variance when all data comes from the same source. As a means of controlling the halo effect, Murphy (1982), compared the overall rating to Spearman’s general factor, or “g.” In that when this factor is removed, a set of specific factors should remain to account for the relationships among the performance items (or dimensions). Murphy criticized this approach as “overkill” (p. 162), i.e., it is too severe, in that both “true” and “error” variance is partialed out by this approach. Hulin (1982) argued that a hierarchical model of human ability would suggest that there is a small, but positive “true” intercorrelation among performance dimensions. Finally, Harvey (1982) criticized the causal ordering that was implied by a partialling approach, i.e., this approach assumes that raters form overall ratings prior to their ratings of specific dimensions. He further recommended structural equation modeling to test competing models of the rating process.

Given the design of our current study, we are in a position to carry out Harvey’s recommendation. Such an approach would, in effect, flip the causal model presented in Model 1, arguing instead that general impression (as captured by overall ratings) causes the ratings of IRB, altruism, and conscientiousness (see Model 3 in Fig. 1). Since this is not our preferred model, we make no specific hypothesis in this regard. However, by using structural equation modeling, we can also test this model to see how well it fits the data. Futornan et al. (1995) employed a similar strategy with structural modeling to test two possible temporal orderings of variables thought to impact student evaluations of their professors.

### 3. Method

#### 3.1. Sample

The data for this study were collected as part of a larger study of employee attitudes developed by the first author.
The sample for this study consisted of the managers of 327 MBA and MS students attending a large Southeastern University during 1992. Manager participation was obtained by asking students for permission to send a short survey of performance behaviors and attitudes to their managers (the employee measures were not included in the current study).

Fig. 1. Three models depicting the ways managers form their views of employees performance.
Of the 327 students, 137 or 41.9% agreed to allow the survey to be mailed to their managers. A total of 101 managers (74%) returned completed surveys resulting in an overall response rate of 31% for managers. The sample of managers was 75% male and 25% female, with a median age of 42, and an average amount of work experience of 9 years. No significant differences were found in comparisons between the sample of employees whose managers did participate in the survey and the sample of employees whose managers did not participate in the survey with regard to age, experience, occupation, or type of organization.

3.2. Measures

One way to classify performance rating scales is by whether they are comparative or absolute. Comparative scales ask raters to compare ratee performance to the performance of other employees, e.g., in a ranking or forced distribution format. Absolute scales, on the other hand, ask raters to compare each ratee against the same standard (or scale values). Examples of absolute scales include Behaviorally Anchored Ratings Scales and Behavioral Observation Scales (Wexley and Yukl, 1984). Both the Smith et al. (1983) and the Williams and Anderson (1991) scales can be classified as absolute scales. In this study, overall employee performance was measured by an eight-item scale. This scale consisted of a seven-item comparative scale, with the addition of a single, global item. The comparative items asked managers to compare employees to their co-workers on seven dimensions of performance (decision-making, problem-solving, work quantity, work quality, initiative, interpersonal skills, and dependability). This scale was coded from 5 = above average to 1 = below average ($\alpha = 0.90$).

In-role performance (IRB) was measured by the seven-item scale ($\alpha = 0.92$) developed by Williams and Anderson (1991). A sample item is, “fulfills responsibilities specified in the job description.” OCB was measured by the 16-item altruism component ($\alpha = 0.86$) and a nine-item conscientiousness component ($\alpha = 0.81$). A sample item from each scale is, “Helps others who have heavy work loads,” and “Does not take extra breaks.” While other dimensions of citizenship have been proposed (Organ, 1988; Graham, 1989), scales measuring altruism and conscientiousness are the most widely used (Organ, 1994; Organ and Ryan, 1995). As noted above, Williams and Anderson’s (1991) OCBI factor was derived from Smith et al.’s (1983) altruism factor. Similarly, the OCBO factor was derived from the conscientiousness factor. Given the aims of our research, we thought it best to use scales that were relatively well-established.

3.3. Analyses

The confirmatory factor analytic procedure for this study was conducted with the PC-version of LISREL VII, with covariance matrices generated by the PRELIS2 pre-processor program (Joreskog and Sorbom, 1989). There were four sets of analyses. The first set of analyses related to examining the distinctiveness of IRBs, altruism, and conscientiousness. Three testlets composed of randomly assigned items were created for each dimension of behavior. Thus, each latent variable in the model had three indicators assuring us of model identification. Using testlets (also called parcels or sub-scales) has the advantage of increasing the subject to indicator ratio without losing the integrity of the construct under examination. This procedure has been used by several authors (Marsh et al., 1989; Shore and Tetrick, 1991; Shore et al., 1995; Williams and Anderson, 1994).

We tested the first hypotheses by fitting two models: (a) a one-factor model with IRB, altruism, and conscientiousness loading on one first-order factor; and (b) a three-factor model, with each measure represented by a unique, but correlated factor. After the measurement model was fitted, a structural equation was incorporated to test Hypotheses 2(a–c). The testlets were retained for analysis in the structural model. IRBs, altruism, and conscientiousness were specified as exogenous latent variables predicting general performance (Model 1 in Fig. 1) that was specified as the single endogenous latent variable. Third, Hypothesis 3 was evaluated with a second-order factor model (Model 2 in Fig. 1), with the IRBs, altruism, conscientiousness, and general performance loading on four first-order factors, and those four first-order factors in turn loading on a general second-order factor. Finally, the structural model was re-specified to evaluate the halo effects model (Model 3 in Fig. 1). This model specified one latent exogenous variable — general performance — predicting three endogenous latent variables, i.e., IRBs, altruism, and conscientiousness.

Model fit to the data was assessed with indices recommended by Medsker et al. (1994) for evaluating multiple models. We included the relative noncentrality index (RNI; McDonald and Marsh, 1990). Medsker et al. noted that the RNI is often more useful than the CFI in model comparison studies, particularly when the RNI is different from the CFI. We also computed the parsimonious normed fit index (PNFI; Mulaik et al., 1989) in order to evaluate the relative parsimony of the various models and the root mean square error of approximation (RMSEA) to evaluate fit relative to the degrees of freedom in the models. The RMSEA represents a close approximation of fit relative to the degrees of freedom when it is between 0.05 and 0.08 (Browne and Cudek, 1993). Chi-square ($\chi^2$) difference tests were employed to evaluated model differences.

4. Results

The means, standard deviations, and reliabilities are presented in Table 1. All correlations among the performance measures were statistically significant. The IRB and general performance scales had the highest correlation ($r = \ldots$)
0.82). The two citizenship dimensions, altruism and conscientiousness, were also highly correlated ($r = 0.56$), similar to results found in other studies (Organ and Konovsky, 1989; Williams and Anderson, 1991; Shore and Wayne, 1993).

The results of the LISREL analyses are presented in Table 2. Our first hypothesis predicted that IRB, altruism, and conscientiousness would represent distinct, yet related aspects of employee performance. The one-factor model with these variables loading on a single factor produced a less than adequate solution ($\chi^2[27] = 147.59$, RNI = 0.795, PNFI = 0.573, RMSEA = 0.21). In contrast, the three-factor model produced a very good fit to the data ($\chi^2[24] = 51.54$, RNI = 0.953, PNFI = 0.614, RMSEA = 0.11). The $\chi^2$ for this model was significantly smaller than for the one-factor model ($\chi^2\text{diff} = 96.05[3], p \leq 0.01$), and both the RNI and PNFI were substantially larger. These results supported our hypothesis that the three dimensions of employee performance represent distinct, yet correlated factors.

Hypotheses 2(a–c) predicted that overall performance would be predicted by IRB, altruism, and conscientiousness. The structural equation results for Model 1 produced a good fit to the data ($\chi^2[48] = 73.60$, RNI = 0.957, PNFI = 0.675, RMSEA = 0.073) and provided partial support for our hypotheses. Both altruism (gamma = $\gamma = 0.34$; $t = 3.86$) and IRBs (gamma = $\gamma = 0.68$; $t = 2.76$) were significant predictors of overall performance. Ninety percent of the variance in overall performance was explained by the structural equation. Contrary to our expectation, conscientiousness was not significantly related to overall performance.

Hypothesis 3 predicted that the four measures of employee performance would be driven by a general, second-order factor of performance. Model 2 in Table 2 depicts this higher-order factor analysis, with the results reflecting a good-fitting model ($\chi^2[50] = 79.84$, RNI = 0.950, PNFI = 0.704, RMSEA = 0.077). The RNI for the second-order model was virtually identical to the RNI for Model 1. Burke et al. (1992) noted that a second-order model will never fit as well as the corresponding first-order model; however, these results indicate that the second-order model yielded a fit that was not significantly different from its corresponding first-order model. In fact, the difference between the models ($\chi^2\text{diff} = 5.74[2]$) was non-significant ($p \leq 0.05$). The second-order loadings from the gamma ($\Gamma$) matrix were high (overall performance = 0.99; IRB = 0.92; altruism = 0.81; and conscientiousness = 0.77), and all were statistically significant ($t > 2.0$).

The indirect effects of the higher-order factor on the four dimensions of performance indicated that the higher-order factor explains a fairly high proportion of variance in the observed variables. The mean total of indirect effects (overall performance, $\mu = 0.67$; IRB, $\mu = 0.62$; altruism, $\mu = 0.50$; conscientiousness, $\mu = 0.48$) support the conclusion that a good proportion of variance in the observed variables is explained by the second-order factor.

Model 3, the “halo” model (overall performance predicting IRBs, altruism, and conscientiousness), resulted in fit indices as follows, $\chi^2[51] = (109.78$, RNI = 0.900, PNFI = 0.681, RMSEA = 0.110). The $\chi^2$ for this model was significantly larger than for Model 1 ($\chi^2\text{diff} = 36.18[3], p < 0.01$). The RNI was lower than in Model 1 and Model 3. The paths from overall performance to IRBs ($\gamma = 0.92$; $t = 5.55$), altruism ($\gamma = 0.80$; $t = 5.74$), and conscientiousness ($\gamma = 0.76$; $t = 5.62$) were all significant and in the expected direction. The structural equation explained 84%, 65%, and 58% of the variance in these three variables, respectively.

### 5. Discussion

The results of this study suggest several important findings for organizational researchers. First, our results replicate the Williams and Anderson (1991) study. The three dimensions of performance (task/IRB, altruism, and conscientiousness) represented three unique correlated dimensions of behavior. This is not an unimportant replication, and provides robust evidence supporting the distinctiveness of OCBs from IRBs (McAllister, 1991). While our study does not directly address the recent debate concerning whether citizenship behaviors are appropriately labeled as “extra role” (Van Dyne et al., 1995; Organ, 1997), we do provide further support for the conceptual frameworks advanced recently by Borman and Motowidlo (1993) and Organ (1997).

### Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2_{\text{null}}$</th>
<th>df_{null}</th>
<th>RNI</th>
<th>PNFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-factor model</td>
<td>147.59</td>
<td>27</td>
<td>-</td>
<td>-</td>
<td>0.795</td>
<td>0.573</td>
<td>0.21</td>
</tr>
<tr>
<td>Three-factor model</td>
<td>51.54</td>
<td>24</td>
<td>96.4*</td>
<td>3</td>
<td>0.953</td>
<td>0.614</td>
<td>0.114</td>
</tr>
<tr>
<td>Structural model 1b</td>
<td>73.61</td>
<td>48</td>
<td>-</td>
<td>-</td>
<td>0.957</td>
<td>0.675</td>
<td>0.073</td>
</tr>
<tr>
<td>Second-order model 2</td>
<td>79.84</td>
<td>50</td>
<td>57.5</td>
<td>5</td>
<td>0.950</td>
<td>0.704</td>
<td>0.077</td>
</tr>
<tr>
<td>Structural model 3d</td>
<td>109.78</td>
<td>51</td>
<td>-</td>
<td>3</td>
<td>0.900</td>
<td>0.681</td>
<td>0.110</td>
</tr>
</tbody>
</table>

* Denotes $\chi^2$ diff test was computed against the one-factor model.
* Denotes structural model with IRB, altruism, and conscientiousness predicting overall performance.
* Denotes $\chi^2$ diff test was computed against the structural Model 1.
* Denotes structural model with overall performance predicting IRB, altruism, and conscientiousness.

### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-role behaviors</td>
<td>4.50 (0.70)</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altruism</td>
<td>4.15 (0.69)</td>
<td>0.62*</td>
<td>0.86*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>4.22 (0.69)</td>
<td>0.66*</td>
<td>0.54*</td>
<td>0.81*</td>
<td></td>
</tr>
<tr>
<td>Rated performance</td>
<td>4.11 (0.71)</td>
<td>0.82*</td>
<td>0.70*</td>
<td>0.63*</td>
<td>0.90*</td>
</tr>
</tbody>
</table>

* $p < 0.05$; reliabilities are on the diagonals. $N = 101$. 
Second, our results replicated and extended work by Podsakoff and MacKenzie (1994), Werner (1994), and others. The dimensions of performance in this study significantly predicted managerial ratings of overall performance. This is consistent with our hypothesis that raters would utilize a “data-driven” or “bottom-up” strategy when processing performance-related information (cf. Lance et al., 1991, 1994b). In our data, IRB and altruism were significant predictors of overall performance, whereas conscientiousness was not. Further, IRB had a stronger influence on overall ratings than did altruism or conscientiousness. This is consistent with earlier studies that have found that IRBs have a stronger impact on overall ratings than do citizenship behaviors (Orr et al., 1989; Werner, 1994). The magnitude of our parameter estimate for altruism is in the same range as those provided by studies cited in Podsakoff et al. (1993). In contrast, the magnitude of our estimate for conscientiousness was considerably below what Podsakoff et al. (1993) reported. It is, however, consistent with recent evidence presented by Van Scotter and Motowidlo (1996).

Given the strength of the first-order correlation between conscientiousness and overall performance, one plausible potential explanation for the lack of significance in our predictive model is that a relatively small sample size led to some problems of multicolinearity, even in the LISREL model. However, in Model 3, the path between general performance and conscientiousness was significant, suggesting that in actuality, once IRBs and altruism were accounted for, no additional explanation of general performance behavior was provided by including conscientiousness. This is the same general finding as was obtained by Van Scotter and Motowidlo (1996) in their study of task and contextual performance among Air Force mechanics, i.e., even though first-order correlations with overall ratings were all large and statistically significant, in their regression analyses, job dedication (conscientiousness) did not account for unique variance in overall ratings. In support of this interpretation, George and Brief (1992) argued against viewing conscientiousness as a citizenship dimension, since indicators of this dimension (e.g., regular attendance, following rules and regulations) are among the in-role requirements of most jobs. In our study as well, it seems that the distinction between ICBs and OCB was least clear when comparing IRB and conscientiousness.

The results of the higher-order factor model supported the “p” general factor of performance forwarded by Vievensvaran et al. (1993). This could potentially raise concerns as to the true distinctiveness of OCB and prescribed role behaviors. Essentially, the higher-order model points out the likelihood that a general understanding of performance may underlie IRBs, altruism, conscientiousness, and overall ratings of performance. As Feldman (1981) pointed out, all evaluations seem to originate from perceptions of ratee traits, thus, it could be that such trait perceptions are having strong effects on all the measures of performance commonly used in organizational research even though those measures incorporate behavioral rather than trait descriptors.

The results were less supportive when we evaluated the halo model, where general impressions (captured by overall performance ratings) were positioned as a precursor of IRB, altruism, and conscientiousness. It must be noted, however, that the halo model fits the data reasonably well, which is consistent with earlier research on the impact of general impressions on ratings of specific performance dimensions (Lance et al., 1991; Lance et al., 1994a,b). While Lance and his colleagues used causal modeling to test three different models of halo, we believe our results serve as the first published test of Hulin’s proposition that researchers use causal modeling to investigate whether general impressions are best viewed as preceding or following ratings of specific dimensions. There is evidence in our data that raters also used a “top-down,” or categorization-based approach to information processing (Lance et al., 1991). Nevertheless, our results suggest that even in the presence of likely halo effects, managers can and do differentiate among IRBs, altruism, and conscientiousness, as suggested by Williams and Anderson (1991).

While the second-order model fits the data somewhat better than did the halo model, it was not possible to directly test the differences between them, since they are not nested models (Lance, 1995). We were also not able to determine the extent that halo remains as a part of this higher-order factor. That is, even though we are arguing that this factor captures “true performance,” some amount of halo is obviously still present. In fact, if statistically partialling to remove halo can be criticized as “overkill” (Murphy, 1982), then our second-order factor approach may appropriately be labeled as “underkill,” in that it understates the impact of halo on these findings.  

Since performance appraisals are used for decision-making within organizations, these results also provide evidence that managers may not limit themselves to information regarding just the in-role performance of employees when making overall performance ratings (MacKenzie et al., 1991). On the positive side of this issue, Waldman (1994) argued that when organizations place strong emphases on teams and total quality management, then the behaviors that have heretofore been classified as citizenship behaviors become critical (i.e., expected) behaviors, in order to ensure the survival and growth of the organization (cf. Morrison, 1994). In such environments, the merging or blurring of in-

---

2 For comparison purposes, we used the same partialling approach utilized by Nathan and Lord (1983). When overall performance was partialled out, the correlations were as follows: $r = 0.06$ for IRB–altruism, $r = 0.30$ for IRB–conscientiousness, and $r = 0.16$ for altruism–conscientiousness, with only IRB–conscientiousness statistically significant at $p < 0.05$. While the implication of this would be that these relationships are virtually all dominated by halo, we agree with Murphy (1982) that the partialling approach is too severe. Evidence favoring this conclusion is provided by the results reported concerning Models 1 and 2.
role and extra-role behaviors may not have as many practical implications. However, organizations need to proceed with caution here. If elements “beyond the job description” (i.e., beyond IRB) are formally utilized in organizational decision-making, this cannot be done in a haphazard or arbitrary manner, since the likelihood of legal challenge in such cases would seem to be high. Future research concerning these issues is definitely needed.

More research is also needed to investigate the prediction of a higher-order factor of performance. This is needed in order to determine the extent of information lost when the higher-order model is employed. Previous research has frequently found that in-role and organizational citizenship dimensions have differential relations with other organizationally relevant outcomes (O’Reilly and Chatman, 1986; Williams and Anderson, 1991; Thompson and Werner, 1997). Motowidlo and Van Scotter (1994), for example, found that IRB and OCB ratings had differential relations with various personal characteristics for a sample of Air Force mechanics. It is our expectation that, in most cases, researchers will want to continue studying these performance dimensions separately, i.e., at the first-order level, so as not to lose the ability to better understand and model these differential relationships. There may be other situations, however, where researchers will want to define performance in this broader, higher-order fashion. One such situation might be in examining the impact of individual performance on team performance issues (Waldman, 1994).

The current study should be replicated with different measures of performance, in order to determine the extent of uniqueness within and across each of these dimensions of employee performance. Recent methodological work on citizenship scales appears promising in this regard (Van Dyne et al., 1994; Moorman and Blakely, 1995). While the current research was restricted to using only the two most commonly used citizenship dimensions, these newer scales also tap other proposed citizenship dimensions (e.g., participation and loyalty). Future research should also include multiple sources of performance ratings, for example, using self- and peer-evaluations, in addition to supervisory ratings (cf. Tornow, 1993 for a discussion of 360° feedback).

Several limitations of our study should be noted. First, the data came from managers in organizations varying in size, type of business, and type of employees, which makes questions of generalizability problematic with regard to the frame of references used to evaluate performance. On the other hand, the dimensions of performance showed a robust nature across individuals in different organizations. Second, the sample size for this study was somewhat small. The ratio of observations to indicators was 8:1, which is less than the desired 10:1 ratio (Williams and Anderson, 1991), but is above the 5:1 ratio cited by Bentler and Chou (1987) as minimally acceptable. Additionally, our sample size exceeded the suggested minimum of 100 for causal modeling (Bollen, 1989). Finally, although we tested models that implied causal ordering, causality cannot of course be determined using cross-sectional data. The use of cross-sectional and same source data always raise questions concerning method bias. In this regard, however, an exploratory factor analysis we conducted did not result in a single method factor accounting for the majority of variance (Harmon, 1967; Podsakoff and Organ, 1986).

In conclusion, this study investigated the extent that four measures of employee performance — comparative performance ratings, IRBs and two dimensions of OCBs (altruism and conscientiousness) — represent unique, yet related constructs. We interpret our results as supporting the argument that managers do distinguish between in-role and extra-role behaviors. However, the results of this study also indicated the existence of a general factor of performance (Viswesvaran et al., 1993). This lends support to researchers such as Waldman (1994), who have argued that both types of employee performance are increasingly critical to organizational success in a rapidly changing business environment. What seems to be called for is a reconceptualization of individual employee performance to encompass both in-role and citizenship performance dimensions (Podsakoff et al., 1993; Organ, 1997). Further research and theorizing along these lines is encouraged (Borman and Motowidlo, 1993; Campbell et al., 1993; Motowidlo and Van Scotter, 1994). In particular, research is needed that can determine the extent that this higher-order factor is capturing “true” performance, vs. general impressions in the minds of raters (Murphy, 1982).

Acknowledgments

This research was supported by an educational grant to the first author from CIBA-Vision. The authors contributed equally to the preparation of this manuscript.

References


MacKenzie SB, Podsakoff PM, Fetter R. Organizational citizenship behavior and objective productivity as determinants of managerial evalua-


