



The relative importance of task, citizenship and counterproductive performance to job performance ratings: Do rater source and team-based culture matter?

Filip Lievens^{1*}, James M. Conway² and Wilfried De Corte¹

¹Ghent University, Belgium

²Central Connecticut State University, CT, USA

This study contributes to our understanding of which factors predict raters' policies for combining performance components into an overall job performance rating. We used a work-roles framework to examine the effects of rater source and team-based culture. The sample consisted of 612 individuals in three job categories (317 nurses, 168 personnel recruiters and 127 sales representatives). Respondents rated employee performance profiles that were described on task, citizenship and counterproductive performance. Raters' weights differed by (a) organizational culture (low- vs. high-team-based); (b) rating source (supervisor vs. peer) and (c) job. In a team-based culture, more weight was given to citizenship performance and less to task performance. Peers attached more importance to citizenship performance and less to task performance. Implications of these findings for performance management are discussed.

The history of criterion research and theory shows a dual focus on unitary, overall criterion measures and measures of performance on separate job dimensions (e.g. Austin & Villanova, 1992; Dunnette, 1963; Schmidt & Kaplan, 1971). Schmidt and Kaplan concluded that both foci are important. Unitary, composite criteria are useful for making decisions (e.g. about personnel selection), whereas dimensional criteria are useful for theoretical understanding of behaviour. Ultimately, increased understanding of the criterion construct should lead to better practice, and theoretical understanding requires a framework that is applicable across specific situations (i.e. across jobs and organizations). Fairly recently, general models of criterion constructs have been proposed and investigated (e.g. Borman & Motowidlo, 1993; Campbell, 1990; Gruys & Sackett, 2003; Organ, 1977). One major construct distinction is that between performance on job-specific tasks (i.e. task performance) and performance of behaviours indicating good organizational citizenship (Bateman & Organ, 1983; Borman

*Correspondence should be addressed to Dr Filip Lievens, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium (e-mail: filip.lievens@ugent.be).

2 *Filip Lievens et al.*

& Motowidlo, 1993). In addition, increased attention has recently been given to another criterion construct, counterproductive performance (Gruys & Sackett, 2003; Rotundo & Sackett, 2002).

The two foci on dimensions and unitary criteria intersect when dimensions must be combined to produce an overall composite criterion, such as a global performance rating. Then, the exact weightings given to different dimensions deserve careful attention because overall job performance might mean something different across raters, if one rater values organizational citizenship behaviours most, whereas another rater attaches most importance to task-related criterion behaviours. Moreover, Murphy and Shiarella (1997) showed that the validity of a selection procedure/battery might vary substantially depending on the different weightings of criterion dimensions (e.g. task and citizenship performance). Thus, it is of key importance to better understand the way raters combine information about different criterion constructs. This might shed light on whether policies are consistent across raters and/or whether policies are a function of personal and organizational characteristics. Rotundo and Sackett's (2002) policy-capturing study showed considerable variability in the weight raters gave to performance dimensions, but did not provide much insight into characteristics that predicted the policies.

In the present study, we used a work-roles framework to examine the effects of rater source and organizational culture on the relative importance attached to three performance components (task, citizenship and counterproductive performance) in determining global job performance ratings. The present study is important because it advances our understanding of the variability found in raters' policies. At a practical level, this study can provide organizations with concrete clues as to how to reduce rater-specific policies and increase organization-specific policies.

Defining task, citizenship and counterproductive performance

One of the major recent developments in criterion theory is the distinction between task and citizenship performance. Task performance has long been recognized as the core of a job and is generally used as a point of comparison (implicitly or explicitly) in describing citizenship performance. Campbell (1990) used the terms job- and non-job-specific task proficiency to denote behaviours and actions needed to produce an organization's goods or services. On the basis of Borman and Motowidlo (1993), we defined task performance as behaviours that (directly or indirectly) contribute to the production of a good or the provision of a service.

Citizenship performance is another criterion construct, originally proposed by Organ as including discretionary behaviour not recognized by formal reward systems (e.g. altruism and civic virtue; Bateman & Organ, 1983; Organ, 1977; Smith, Organ, & Near, 1983). Borman and Motowidlo (1993) proposed a similar construct which they labelled contextual performance. Since then, Organ (1997) and Coleman and Borman (2000) have moved towards merging the concepts. They have noted problems with defining citizenship as discretionary, hence our definition, based on Borman and Motowidlo is the behaviour that contributes to the goals of the organization by contributing to its social and psychological environment.

Along with task and citizenship performance, another important criterion construct is counterproductive performance. Counterproductive work behaviours consist of a broad array of behaviours that violate the organization's legitimate interests, including among others theft, unsafe behaviour and misuse of information, time or resources

(Sackett & DeVore, 2001). In the current study, we used Rotundo and Sackett's (2002) definition of counterproductive performance: 'voluntary behavior that harms the well-being of the organization' (p. 69).

Previous research on weighting of task, citizenship and counterproductive performance

Interest in how raters weight criterion dimensions when combining them into overall ratings of job performance has a long research tradition (e.g. Naylor & Wherry, 1965). While valuable, these early studies understandably did not make the distinction between citizenship, task and counterproductive performance. More recent studies have included task and citizenship performance but not counterproductive performance, and they have consistently shown that each performance construct receives significant weight. MacKenzie, Podsakoff, and Fetter (1991, 1993) regressed managers' ratings of salespeople's performance onto objective production scores and ratings of citizenship. They found roughly equal weight (though it is possible that the weight for citizenship was inflated relative to production by same-source bias in citizenship and overall ratings). Both Motowidlo and Van Scotter (1994) and Van Scotter and Motowidlo (1996) showed that measures of task and contextual (i.e. citizenship) performance independently contributed to supervisors' overall performance ratings, and that task performance showed greater incremental prediction over contextual performance than vice versa (indicating greater weight for task performance). Conway (1999) concluded that supervisor raters gave more weight to technical than to contextual performance (i.e. dedication and interpersonal performance), but that peer raters gave roughly equal weights. Finally, Johnson (2001) found that across a variety of job families, supervisor raters gave significant and about equal weight to task proficiency and citizenship.

A key limitation of all the studies cited in the previous paragraph is that they conducted analyses across raters, and thus differences among raters could not be assessed. These across-rater analyses from previous research focus on differences between raters, which could include not only 'real' differences in ratees' performance but also different rating tendencies (e.g. leniency or severity bias). As it is difficult to point out the effect the rating tendencies have had on results, it would be desirable to conduct analyses that rule out this source of error. In addition, across-rater analyses do not allow studying differences in raters' judgment policies (e.g. one cannot see how much individual raters differ in the weight given to citizenship performance).

A policy-capturing approach overcomes these limitations, and this within-rater approach was used by Rotundo and Sackett (2002). They asked managers in five jobs to make overall ratings of performance based on written profiles of hypothetical workers. Each profile included information about the worker's task, citizenship and counterproductive performance, and each manager rated 34 profiles. Within-rater regression analyses were used to identify each rater's policy. Rotundo and Sackett's study revealed that, overall, the most weight was given to task performance (positive weight) and counterproductive performance (negative weight), with less weight given to citizenship performance (positive weight). On the basis of their results and the other aforementioned studies, we expected all the three performance constructs to contribute to overall performance ratings. This leads to Hypothesis 1, which attempts to replicate prior findings.

Hypothesis 1: Task, citizenship and counterproductive performance will all be given significant weight in ratings of overall performance.

Factors affecting raters' policies

Rotundo and Sackett also discovered that raters varied considerably in the amount of weight given to each performance construct. This result prompts the question as to which factors might explain this variation. To identify possible factors that contribute to the variability in raters' policies, this study uses Katz and Kahn's (1978) concept of work-roles as a theoretical framework. According to Katz and Kahn, 'roles describe specific forms of behavior associated with given positions' (p. 43). The position of emergency room nurse, for example, has certain behaviours associated with it. Given the interdependent nature of organizations, a worker in a given role is linked to workers in other roles (e.g. emergency room physicians; nursing supervisor; other nurses) and collectively, these workers form a role set. Understanding role behaviour requires analysing the social context and identifying the 'role expectations', or beliefs and attitudes about what someone in a particular role should do. For example, a nurse is depended on by other nurses, by the nursing supervisor, etc., so the other nurses and the supervisor have expectations for what that nurse should be doing.

Role expectations are mental constructs and expectations for a given worker and may differ across other members of the role set. Katz and Kahn suggested, for example, that supervisors and peers may have different ideas about what a worker should do. This suggests that *rater source* is a first factor that might determine role expectations. Similarly, role expectations as mental constructs might be shaped by the specific context wherein people work. Especially relevant might be whether people do or do not work in a team-based culture because this will create an even greater interdependency among the various roles. This indicates that *team-based organizational culture* might also be a key determinant of role expectations.

Performance evaluations (e.g. by a supervisor or a peer) can be thought of as assessments of how well a worker has conformed to a rater's role expectations. These evaluations may therefore differ across raters who have different role expectations, even if the raters observe the same behaviour. In-turn, the two factors (rater source and team-based culture) theoretically linked to role expectations might also influence these performance evaluations. Below, we develop these general ideas further into testable hypotheses by relating them to the importance attached to different performance components.

Rating source

Comparing different rating sources (especially supervisors and peers) has been a frequent research topic (see Borman, 1997). For example, both Harris and Schaubroeck (1988) and Conway and Huffcutt (1997) found relatively low correlations between supervisor and peer raters. Borman suggested several reasons for the less-than-perfect agreement. The possibilities include: (a) observation of different behaviours by supervisors vs. peers (i.e. so-called ecological perspective); (b) use of different performance dimensions by supervisors vs. peers and (c) use of different weights by supervisors vs. peers when combining dimensions into overall ratings. The latter possibility would occur if supervisors and peers have different role expectations. We tested this idea by comparing the policies of supervisors and peers to examine differences in their weighting of task, citizenship and counterproductive performance.

Supervisors' role expectations include behaviours leading to productivity by their organizational units, and this includes good task performance from the supervisors' subordinates (Conway & Huffcutt, 1997). Supervisors probably do not see as clear a link between subordinates' citizenship performance and productivity as they do between

task performance and productivity. Peer raters' role expectations, on the other hand, will probably focus more on working cooperatively. For peers, ratees that are good citizens of the organization (e.g. being helpful and cooperative) should be valued more than they might be by supervisors. In other words, this leads to the prediction that supervisors will place more weight on task performance than will peers, while peers will place more weight on citizenship performance than will supervisors.

There is some empirical evidence to support our theoretical assertion. Borman, White, and Dorsey (1985) found that supervisor raters had a higher path coefficient for a hands-on technical proficiency measure than for ratings of interpersonal factors, while peer raters showed the opposite pattern. And as noted earlier, Conway (1999) found that supervisor raters gave more weight to technical than to contextual performance, but that peer raters gave roughly equal weight. Please note that while these studies provided important evidence on peer vs. supervisor weighting of information, they conducted analyses across raters and did not include counterproductive performance (as we do in the present study).

Role expectations can also be applied to counterproductive performance. However, we do not hypothesize any supervisor-peer differences in its weighting. Counterproductive performance should be contrary to expectations of both supervisors (making it difficult to achieve work objectives) and peers (making it difficult for peers to get their jobs done). In fact, there is evidence of a 'negativity bias' in which raters give more weight to evidence of poor performance (e.g. Ganzach, 1995; Skowronski & Carlston, 1989). We can therefore imagine both sources giving substantial weight to counterproductivity. Hence, we offer hypotheses regarding supervisor-peer differences only about task and citizenship performance.

Hypothesis 2a: The relative importance of task to global job performance ratings will be significantly higher when supervisors are raters as compared with peers.

Hypothesis 2b: The relative importance of citizenship to global job performance ratings will be significantly higher when peers are raters as compared with supervisors.

Team-based organizational culture

On the basis of Katz and Kahn's (1978) concept of role expectations, it can also be assumed that a team-based culture will influence organization members' (such as raters') role expectations and the roles that people play. It has been noted that citizenship behaviours can be seen as in-role or as extra-role (e.g. Vey & Campbell, 2004). We believe that citizenship behaviours will be more likely to be included in role expectations in an organization with a team-based culture. In such an organization, a premium would be placed on cooperating with others, especially team members, and more generally on promoting the social and the psychological environment necessary for effective team performance. If citizenship is considered part of the role expectation in a team-based organization, then citizenship should receive more weight in determining overall performance than it would in a traditional organization.

Similar assertions have been made in prior research. For example, Werner (2000) speculated that citizenship behaviour would be facilitated by a team-based culture. Further, Johnson (2001) suggested that in team-based organizations, citizenship would get more weight in overall ratings than it would in organizations that are not team-based, whereas task performance would get more weight in more traditional organizations. However, empirical evidence is still lacking.

6 Filip Lievens et al.

As with rating source differences, we believe role theory is relevant to team-based culture and counterproductivity. We speculate that counterproductive performance will be given more (negative) weight in a team-based organization because counterproductivity will be perceived as a breach of the role expectations that should develop in a team-based environment.

Although a team-based culture is essentially not intended to deemphasize the importance of task performance as its purpose is to emphasize team work when completing core tasks, we still speculate that task performance will play a less dominant role in raters' agenda due to the increased relative importance of citizenship and counterproductive performance. All these lead to our third set of hypotheses.

Hypothesis 3a: The relative importance of citizenship to global job performance ratings will be significantly higher in organizations with a team-based organizational culture.

Hypothesis 3b: The relative importance of counterproductive to global job performance ratings will be significantly higher in organizations with a team-based organizational culture.

Hypothesis 3c: The relative importance of task to global job performance ratings will be significantly lower in organizations with a team-based culture.

Method

Sample

This study was conducted in the Flemish part of Belgium. To increase generalizability, we conducted the study in three different jobs: nurses, personnel recruiters and sales representatives. These jobs were chosen because they represented a diverse set of jobs. To obtain the participation of incumbents of these three jobs, we started by gathering an extensive list of organizations that employed individuals in these jobs. Next, a random set of organizations within each geographical area was contacted by research assistants via e-mail or by telephone. If the organization agreed to participate, we sent questionnaires accompanied by a reference letter to the contact person in the organization who in-turn was responsible for distributing these materials further. Respondents received the questionnaires in person at work. Following this procedure, questionnaires were distributed to a total of 1206 individuals.

Only people in the three job categories (either working as employees or supervisors) were asked to voluntarily complete the questionnaire. Our inclusion criteria were based on the O*NET descriptions of these jobs. For the nurses, only the O*NET category 'Registered Nurses' working in hospitals was included. These nurses typically work under close supervision in a structured environment. For the recruiters, the O*NET category 'Personnel Recruiters' was the focus. In particular, we asked only the participation of recruiters working in either general or sector-specific recruitment agencies. For the sales representatives, we focused on the O*NET category 'Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products'.

We received complete and usable data from 612 individuals, consisting of 317 nurses (217 females, 100 males; average age = 37.7 years; average work experience = 15.6 years, response rate of 63%), 168 personnel recruiters (149 females, 19 males; average age = 30.7 years; average work experience = 8.3 years, response rate of 42%) and 127 sales representatives (37 females, 90 males; average age = 39.5 years; average work experience = 17.2 years, response rate of 43%). If possible, we checked whether the

individuals in these three categories were representative of the general population. For example, national reports confirmed that the gender and the age of our sample of nurses were similar to those characteristics in the nursing population (Pacolet, Van De Putte, Cattaert, & Coudron, 2002). However, no such reports were available for the other two job categories.

Development of employee performance profiles

We constructed employee performance profiles that were described on task, citizenship and counterproductive performance. The construction of the performance profiles proceeded through four stages (see Rotundo & Sackett, 2002). In the first stage, a representative set of behaviours per performance component was gathered. Per job, behaviours for task performance were retrieved from O*NET and job analyses. We identified counterproductive and citizenship performance behaviours by inspecting overview articles and measures. This process of gathering representative sets of behaviours produced a list of 8–10 behaviours per performance component. We then constructed low, medium and high examples of each behaviour, resulting in a list of 24–30 behaviours per component.

Second, we determined the performance level of the behaviours to guarantee distributional equivalence. Pre-studies in each occupational group (37 nurses/nursing managers, 34 personnel recruiters/managers and 24 sales representatives/sales managers) were conducted. After reading a definition of each performance component, respondents used seven-point scales to rate the task behaviours from *low task* to *high task performance*, the citizenship behaviours from *low citizenship* to *high citizenship performance* and the counterproductive behaviours from *low counterproductive* to *high counterproductive performance*. Next, we computed descriptive statistics (see Table 1) per behaviour and selected a final set of behaviours that included behaviours: (1) with a standard deviation below 1.5 (indicative of sufficient agreement); (2) with a similar mean and variance per performance component (indicative of distributional equivalence) and (3) at the low, medium and high end of the effectiveness continuum (indicative of a normal distribution). Twelve behaviours per performance component survived these inclusion criteria.

Table 1. Descriptive statistics for the study behaviours per performance component

	Task performance		Citizenship performance		Counterproductive performance	
	M	SD	M	SD	M	SD
Nurse	3.78	1.21	3.79	1.17	3.78	1.16
Personnel recruiter	3.79	1.22	3.79	1.17	3.78	1.16
Sales representative	3.78	1.22	3.79	1.17	3.78	1.16

We also examined the realism of the behaviours in one of the occupational groups of the main study. Twenty-one nurses (16 females, 5 males; mean age = 39.2 years; mean working experience = 11 years) rated each of the statements on a seven-point scale, ranging from 1 = *not realistic at all* to 7 = *very realistic*. The average realism rating was 4.66 ($SD = 1.83$). There were four behavioural statements with a realism rating

slightly below 3.5. Each of these four behaviours was situated at the negative end of the effectiveness continuum. Although these four behaviours might be considered somewhat less realistic, we included them because we wanted to sample behaviours spanning the whole effectiveness continuum.

The third stage consisted of developing the performance profiles. To balance concern over sampling error with concern for participant fatigue (Cooksey, 1996), we chose a cue-profile ratio of 1 to 8, yielding 24 profiles. As it is important that the behaviours gathered in stage 2 are combined in a realistic way, we ensured that the correlation among the performance components built into the profiles simulated values found for real people (Borman, 1978). To this end, we used the best available estimates. On the basis of the recent large-scale review of Viswesvaran, Schmidt, and Ones (2005; see also Sackett & DeVore, 2001), the correlation between task and counterproductive performance was set at $-.55$ and the correlation between counterproductive and citizenship performance at $-.60$. On the basis of the same large-scale review (Tables 4 and 5 in Viswesvaran *et al.*, 2005), the correlation between task and citizenship performance was set at $.55$, although we acknowledge there exists some debate about this value (see Motowidlo & Van Scotter, 1994).

Fourth, we randomized the order in which information about the performance components was presented in the profiles to eliminate possible primacy and recency effects. The final set consisted of 26 profiles (the 24 profiles and 2 repeat profiles to assess reliability). The following is an example profile for the job of nurses: 'Karen accurately conducts physical examinations and hereby systematically follows existing procedural guidelines. She never steals material from the hospital. She sometimes speaks positively about the hospital to other nurses or patients.'

Questionnaire

After the study introduction, the questionnaire contained information about the organization and the role of the employee in the organization (see Experimental manipulations). Next, participants were asked to read each profile at the time and to use only the profile information for rating the job incumbent on a five-point Likert scale: 1, *low overall job performance* to, 5, *high overall job performance* (see Rotundo & Sackett, 2002). Finally, a manipulation check (see below) and demographic questions were included.

Experimental manipulations

The overall design was a 2 (team-based culture) \times 2 (rater source) between-subjects design. We did not decide to use participants' self-reports of the actual culture of their organizations because as such it would be virtually impossible to examine the effects of culture without bringing in all kinds of confounding factors. Instead, we chose to manipulate team-based culture by developing two different portrayals of organizations (cf. Schleicher & Day, 1998). Organization A was portrayed by being high on team orientation. Conversely, Organization B was characterized as an organization low on team orientation. We tried to cast and evoke a realistic and vivid portrayal of these organizational cultures by including pictures and behavioural incidents. For example, a picture of people happily working together evoked the team-based Organization A. The behavioural incidents were gathered from actual employees working in the three jobs and from organizational culture frameworks. For example, O'Reilly, Chatman, and Caldwell

(1991) found that the following attributes loaded on team orientation: collaboration, team-oriented and people-oriented. This resulted in the following behaviours for Organization A: 'employees work constructively together in teams and help each other', 'there does not exist a large distance between employees and supervisors (e.g. supervisors and employees often meet after work)' and 'the organization values participative decision making'. The common thread running through these behaviours is that people do things (executing tasks, taking decisions or time after work) together. Organization B was depicted with the reverse behavioural incidents.

To assess the strength of this manipulation, we randomly distributed the organizational descriptions to 27 master's I/O psychology students (19 females, 8 males; mean age = 20.7 years) who rated them on two items ($\alpha = .76$). The anchors of the first item ranged from 1 = *not at all team-based* to 7 = *very team-based*, whereas the anchors of the second item varied from 1 = *not at all people-oriented* to 7 = *very people-oriented*. Participants rated Organization A ($M = 5.35$, $SD = 0.68$) as significantly higher on team orientation as Organization B ($M = 1.67$, $SD = 0.44$), $t(27) = 16.4$, $p < .01$, $d = 1.9$. So, the manipulation was salient.

The second factor, rater source, had two levels (colleague vs. supervisor). This factor was operationalized by mentioning that the participant worked either as a supervisor or as a colleague. To ensure that respondents' experience level matched the judgment task (Aiman-Smith, Scullen, & Barr, 2002), only the individuals working as supervisors received a questionnaire with supervisory role instructions. Similarly, only individuals working as colleagues received a questionnaire with colleague role instructions. Within the two groups (colleagues and supervisors), questionnaires were randomly distributed.

Manipulation check

After rating all employee performance profiles, we asked participants to indicate which organizational culture they had in mind when rating the employee profiles. To this end, the same two items as above ('team-based' and 'people-oriented'; $\alpha = .76$) were used. Participants assigned to the team-based organizational culture condition reported the organizational culture that they used to rate the profiles to be significantly more team-based ($M = 5.75$, $SD = 0.74$) than participants assigned to the opposite condition ($M = 2.25$, $SD = 0.62$), $t(610) = -63.65$, $p < .01$, $d = 1.6$. These manipulation check results showed that participants had indeed used the appropriate organizational information when rating the profiles.

Analyses

Regression analyses are typically conducted to capture people's decision policy. However, when intercorrelation among predictors exists (as is the case in this study), regression coefficients have long been judged inadequate to indicate the relative importance of a predictor because the impact of one predictor cannot be considered when holding the other predictors constant (Budescu, 1993; Hoffman, 1962). Currently, there are two preferred methods for determining a predictor's relative importance: Budescu's dominance analysis and Johnson's (2000, 2001) relative weights. According to Johnson and LeBreton (2004), both indices take a predictor's direct effect and its effect when combined with other predictors into account, and both yield importance weights that represent the proportionate contribution each predictor makes to R^2 . When they are used for analysing the same data, both indices produce virtually the same results. Here,

we computed Johnson's relative weights per rater (expressed as proportions of R^2) because they are easier to compute than Budescu's dominance analysis.

Results

Preliminary analyses

As two repeat profiles were included, we could examine the within-rater reliability of the ratings made. Across all raters and jobs, the overall reliability coefficient for the three repeat profiles was .86, indicating that raters used a consistent policy when making ratings.

Test of hypotheses

Table 2 presents means and standard deviations of the relative importance of the three performance components to global job performance ratings per job and across jobs. All three performance components accounted for substantial portions of the variance, confirming that they all matter when making global job performance ratings. Across the three jobs, task performance explained 37% of the variance, citizenship performance 33% and counterproductive performance 30%. These results support Hypothesis 1.

Table 2. Means and standard deviations of the relative importance weights of performance components broken down by job

	<i>M</i>	<i>SD</i>	<i>N</i>
Nurses			
Task performance	0.44	0.09	317
Citizenship performance	0.28	0.07	317
Counterproductive performance	0.28	0.07	317
Personnel recruiters			
Task performance	0.24	0.07	168
Citizenship performance	0.40	0.08	168
Counterproductive performance	0.36	0.08	168
Sales representatives			
Task performance	0.39	0.09	127
Citizenship performance	0.33	0.08	127
Counterproductive performance	0.28	0.08	127
All jobs			
Task performance	0.37	0.12	612
Citizenship performance	0.33	0.09	612
Counterproductive performance	0.30	0.09	612

Note. The relative importance weights were computed on the basis of the procedure of Johnson (2000, 2001).

Next, we conducted a MANOVA with the relative weights for each of the three performance components as dependent variables and with job, rater source and team-based culture as independent variables. There were multivariate main effects for job, rater source and team-based culture. There were no significant interaction effects. Below each of the main effects is described.

First, the MANOVA showed a main effect for job, $F(6, 1196) = 92.05$, $p < .001$, partial $\eta^2 = .32$. Follow-up ANOVAs indicated that the type of job had significant effects

on all three performance components. Task performance received the most weight in the nurse and the sales representative jobs. Interestingly, citizenship performance was given the most weight in the job of personnel recruiter.

Second, there was a significant multivariate main effect for rater source, $F(3, 598) = 11.55, p < .001$, partial $\eta^2 = .06$. Follow-up ANOVAs showed that rater source had a significant effect on the relative importance attached to task performance, $F(1, 600) = 22.44, p < .001$, partial $\eta^2 = .04$, and on the relative importance attached to citizenship performance, $F(1, 600) = 28.10, p < .001$, partial $\eta^2 = .05$. Table 3 presents the means and standard deviations of the importance weights, broken down by rater source. In line with Hypothesis 2a, the relative importance of task to global job performance ratings was significantly higher among supervisor raters ($M = 0.39, SD = 0.12$) than among peer raters ($M = 0.36, SD = 0.12$). In line with Hypothesis 2b, the importance given to citizenship performance was significantly higher among peer raters ($M = 0.34, SD = 0.09$) than among supervisor raters ($M = 0.31, SD = 0.08$).

Table 3. Means and standard deviations of the relative importance weights of performance components broken down by rater source

	Supervisors			Peers		
	M	SD	N	M	SD	N
Nurses						
Task performance	0.44 _a	0.09	164	0.44 _a	0.09	153
Citizenship performance	0.27 _a	0.06	164	0.30 _b	0.07	153
Counterproductive performance	0.29 _a	0.07	164	0.27 _b	0.08	153
Personnel recruiters						
Task performance	0.25 _a	0.07	75	0.22 _b	0.06	93
Citizenship performance	0.39 _a	0.07	75	0.41 _a	0.07	93
Counterproductive performance	0.36 _a	0.07	75	0.36 _a	0.08	93
Sales representatives						
Task performance	0.42 _a	0.10	59	0.36 _b	0.07	68
Citizenship performance	0.31 _a	0.08	59	0.35 _b	0.07	68
Counterproductive performance	0.28 _a	0.08	59	0.29 _a	0.08	68
All jobs						
Task performance	0.39 _a	0.12	298	0.36 _b	0.12	314
Citizenship performance	0.31 _a	0.08	298	0.34 _b	0.09	314
Counterproductive performance	0.30 _a	0.08	298	0.30 _a	0.09	314

Note. For the relative weights, see note to Table 2. Only values in the same row with a different subscript differ significantly ($p < .05$) from each other across conditions.

Table 3 also shows that the effects were relatively robust across jobs. Yet, in the job of nurses, there was an additional finding, namely a significant effect of rater source on the importance given to counterproductive performance, $F(1, 315) = 10.54, p < .01$, partial $\eta^2 = .03$. Nursing managers attached more importance to counterproductive performance than nurses.

Third, the MANOVA showed a significant multivariate main effect for team-based culture, $F(3, 598) = 12.84, p < .001$, partial $\eta^2 = .06$. Follow-up ANOVAs showed that team-based culture had a significant effect ($p < .001$) on the relative importance attached to task performance, $F(1, 600) = 20.22, p < .01$, partial $\eta^2 = .03$, and citizenship

performance, $F(1, 600) = 34.88, p < .001$, partial $\eta^2 = .06$. Table 4 presents the descriptive statistics of the importance weights broken down by team-based culture. In line with Hypothesis 3a, the importance attached to citizenship performance was higher in more team-based ($M = 0.34, SD = 0.08$) than in less team-based organizations ($M = 0.31, SD = 0.09$). Consistent with Hypothesis 3c, the importance attached to task performance was lower in more team-based ($M = 0.36, SD = 0.11$) than in less team-based organizations ($M = 0.39, SD = 0.13$). The effect on counterproductive performance was not significant, lending no support to Hypothesis 3b.

Table 4. Means and standard deviations of the relative importance weights of performance components broken down by organizational culture

	Low team-based			High team-based		
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>
Nurses						
Task performance	0.46 _a	0.09	158	0.42 _b	0.09	159
Citizenship performance	0.27 _a	0.07	158	0.30 _b	0.06	159
Counterproductive performance	0.27 _a	0.08	158	0.28 _a	0.07	159
Personnel recruiters						
Task performance	0.24 _a	0.06	83	0.23 _a	0.06	85
Citizenship performance	0.38 _a	0.07	83	0.42 _b	0.07	85
Counterproductive performance	0.37 _a	0.08	83	0.35 _a	0.07	85
Sales representatives						
Task performance	0.41 _a	0.10	63	0.36 _b	0.08	64
Citizenship performance	0.31 _a	0.08	63	0.35 _b	0.07	64
Counterproductive performance	0.28 _a	0.09	63	0.29 _a	0.08	64
All jobs						
Task performance	0.39 _a	0.13	304	0.36 _b	0.11	308
Citizenship performance	0.31 _a	0.09	304	0.34 _b	0.08	308
Counterproductive performance	0.30 _a	0.09	304	0.30 _a	0.08	308

Note. For the relative weights, see note to Table 2. Only values in the same row with a different subscript differ significantly ($p < .05$) from each other across conditions.

Additional analyses

As our central purpose was to better understand the variability in raters' weights, we examined other factors that might influence these weights. We investigated the effects of various demographic variables (gender, educational level and age). As the jobs differed in terms of demographic characteristics, the analyses were done per job. In some jobs, the demographic variables had small albeit significant effects. Among sales representatives, a MANOVA with educational level (with or without university degree) as independent variable and the performance components as dependent variables showed a significant effect for educational level, $F(3, 119) = 3.43, p < .05$, partial $\eta^2 = .08$. Follow-up analyses indicated that this effect was due to task performance, $F(1, 121) = 9.70, p < .01$, partial $\eta^2 = .07$, and counterproductive performance, $F(1, 121) = 5.24, p < .05$, partial $\eta^2 = .04$. People with a university degree attached more importance to task performance ($M = 0.42, SD = 0.11$) and less to counterproductive performance ($M = 0.26, SD = 0.09$) than people without a university degree ($M = 0.37, SD = 0.08$ and $M = 0.30, SD = 0.08$, respectively). Among nurses,

there was a significant effect for age, $F(3, 313) = 3.55, p < .05$, partial $\eta^2 = .03$. This effect was due to counterproductive performance, $F(1, 315) = 7.75, p < .01$, partial $\eta^2 = .02$. Older people ($M = 0.29, SD = 0.07$) attached more importance to this component than younger people ($M = 0.27, SD = 0.07$). Gender effects were observed in none of the jobs.

Discussion

Main contributions

This study built on work-role theory to increase our understanding of potential factors that might explain the variability among raters in the weights given to different criterion constructs. To this end, within-rater analyses were conducted. Most of our hypotheses were confirmed. A first factor uncovered was that the weights differed by rating source (supervisor vs. peer). We found that peers placed more weight on citizenship than did supervisors (Borman *et al.*, 1985; Conway, 1999). In addition, there was evidence of source differences in weighting counterproductive performance. For nurses (though not for the other jobs), supervisors seemed to give more weight than peers to counterproductive performance. Second, team-based culture explained some of the variability in raters' policies for combining the performance components into an overall job performance rating. This second factor provides context for work behaviour, including performance-rating behaviour. We found that in a team-based culture, more weight was given to citizenship performance and less to task performance. While this type of effect has been predicted (Johnson, 2001), we believe we are the first to demonstrate it.

Although the effects for rater and organizational culture were statistically significant and relatively robust across three jobs, these effects were not large. The effect sizes associated with the two manipulated factors were typically about .05. The effects of additional rater factors (e.g. age) were equally small. Our finding that both the effect sizes of manipulated factors and natural variables such as age are rather small might indicate two things. On one hand, it might indicate that these rater-based or organization-based effects on raters' policies are indeed small. If this is correct, then raters' policies might have more modest effects on the meaning and validity of overall job performance across organizations than previously thought (Murphy & Shiarella, 1997). On the other hand, the small effects found might also suggest that we have only started to understand the variability in raters' policies.

In this study, we focused on factors that could be theoretically linked to Katz and Kahn's (1979) work-role theory. Yet, other potential moderators should be examined. One fruitful avenue consists of linking raters' personality traits to their policies of weighting performance components. For example, it seems likely that people high on agreeableness or need for affiliation will attach more importance to citizenship behaviour than people low on these traits. Similarly, raters' implicit leadership theories (e.g. task vs. relationship orientation; Engle & Lord, 1997) might explain some of the variability in raters' policies.

Implications for performance appraisal

What do these results mean for performance appraisal? One implication of our source-differences findings is a greater understanding of the low-to-moderate correlations between sources on overall ratings (e.g. Conway & Huffcutt, 1997). The low or moderate relationships are probably due, in part, to differences between individual raters regardless of source (as indicated by individual variability in our participants'

policies), but are probably also due, in part, to differences in sources' weightings of task, citizenship and counterproductive performance (Borman, 1997). This might be taken into account when interpreting overall ratings in multisource feedback - disagreement between sources could be a function of differing weights rather than differing evaluations of behaviour.

An implication of our team-based culture effects is that raters might be capable of matching their rating policies to an organization's culture. This is important because according to strategic HRM (Fombrun, Tichey, & Devanna, 1984) organizations may wish to impose a performance theory so that the organization's predetermined policy replaces to a certain extent raters' implicit policies and both policies are aligned. But the success of this so-called 'horizontal alignment' depends on raters being able to modify their judgment policies - and our research suggests they can *to some extent*.

A final implication for performance appraisal deals with rater training. Whereas rater training approaches such as frame-of-reference training include a discussion of organizational norms (Schleicher & Day, 1998; Woehr & Huffcutt, 1994), this represents only one side of the equation. The variability in raters' policies shows that raters' idiosyncratic rating policies also deserve attention. Attention should be paid to the organization-specific weights for combining different performance components into an overall rating. It is vital to impose not only the organization-specific behaviours but also the organization-specific performance standards to raters (see content of frame-of-reference training approaches). In addition, it seems pivotal to familiarize raters with how to integrate the various dimensions into an overall performance rating (in case one does not use a formula to arrive at this rating).

Whereas the effects associated with rater source and organizational culture were rather small, we found clear evidence that weightings were different across jobs (effect size of .34). Task performance was given the most weight for nurses (considerably more than the weight given to citizenship or counterproductive performance) and the least weight for personnel recruiters (considerably *less* than for citizenship or counterproductive performance). In addition, citizenship performance was even the most important component among personnel recruiters. We can only speculate on the reasons, but the core tasks involved in personnel recruiting are likely to be more interpersonal than are those of nursing. An occupation search on O*NET (Peterson, Mumford, Borman, Jeanneret, & Fleisman, 1999) showed that the first two tasks listed for personnel recruiters were 'Establish and maintain relationships with hiring managers. . . .' and 'Interview applicants to obtain information. . . .' These tasks are more interpersonal in nature than the first two listed for registered nurses: 'Maintain accurate, detailed reports and records' and 'Monitor, record and report symptoms and changes in patients' conditions'. Hence, it seems intuitive that more weight would be given to task performance for nurses, while more weight is given to citizenship performance for personnel recruiters. More systematic evidence could be provided by relating raters' weights to measures of occupational characteristics for a variety of jobs.

Limitations

This study was an experimental policy-capturing study that typically maximizes internal validity (i.e. systematic sampling of stimuli, control over confounding) at the expense of external validity (Aiman-Smith *et al.*, 2002; Karren & Barringer, 2002). Therefore, we tried to increase the external validity of our study using a field sample and three jobs. We also ensured that the intercorrelation among performance components was consistent

with the best available estimates. Yet, other aspects were less realistic as short written profiles and organizational descriptions were used. In addition, no strong evidence was available about the representativeness of the three samples of this study. Still, our results were consistent with field research on some points (e.g. greater weight to task performance by supervisors and to citizenship performance by peers). Whether our other results generalize to field settings is a question for future field research.

Conclusions

This study showed that the variability in raters' policies to combine different performance components into an overall job performance rating should be given more attention. In particular, we need research that examines factors that predict raters' policies. This study constitutes a first step in this line of research by documenting the impact that rater source and team-based culture might have on raters' policies. Future studies are needed to investigate other influencing factors such as people's personality traits and implicit leadership theories.

Acknowledgements

We would like to thank Veerle Van Geluwe for her help in collecting the data.

References

- Aiman-Smith, L., Scullen, S., & Barr, S. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy- capturing and other regression-based techniques. *Organizational Research Methods*, 5, 388-414.
- Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, 77, 836-874.
- Bateman, T. S., & Organ, D. W. (1983). Job satisfaction and the good soldier: The relationship between affect and employee citizenship. *Academy of Management Journal*, 26, 587-595.
- Bennett, R. J., & Robinson, S. L. (2000). The development of a measure of workplace deviance. *Journal of Applied Psychology*, 85, 349-360.
- Borman, W. C. (1978). Exploring the upper limits of reliability and validity in job performance ratings. *Journal of Applied Psychology*, 63, 135-144.
- Borman, W. C. (1997). 360 degree ratings: An analysis of assumptions and a research agenda for evaluating their validity. *Human Resource Management Review*, 7, 299-315.
- Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt, W. C. Borman, & Associates (Eds.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.
- Borman, W. C., White, L. A., & Dorsey, D. W. (1985). Effects of rater task performance and interpersonal factors on supervisor and peer performance ratings. *Journal of Applied Psychology*, 80, 168-177.
- Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114, 542-551.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial & organizational psychology* (2nd ed., Vol. 1, pp. 687-732).
- Coleman, V. I., & Borman, W. C. (2000). Investigating the underlying structure of the citizenship performance domain. *Human Resources Management Review*, 10, 25-44.
- Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology*, 84, 3-13.

- Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multi-source performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*, 331-360.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic press.
- Cooper, W. H., & Richardson, A. (1986). Unfair comparisons. *Journal of Applied Psychology, 71*, 179-184.
- Q4 Dunnette, M. D. (1963). A note on the criterion. *Journal of Applied Psychology, 47*, 251-253.
- Engle, E. M., & Lord, R. G. (1997). Implicit theories, self-schemas, and leader-member exchange. *Academy of Management Journal, 40*, 988-1010.
- Fombrun, C., Tichay, N. M., & Devanna, M. A. (1984). *Strategic human resource management*. New York: Wiley.
- Ganzach, Y. (1995). Negativity (and positivity) in performance evaluation: Three field studies. *Journal of Applied Psychology, 80*, 491-499.
- George, J. M., & Jones, G. R. (1997). Organizational spontaneity in context. *Human Performance, 10*, 153-170.
- Q4 Gruys, M. L., & Sackett, P. R. (2003). Investigating the dimensionality of counterproductive work behavior. *International Journal of Selection and Assessment, 11*, 30-42.
- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.
- Hoffman, P. J. (1962). Assessment of the independent contributions of predictors. *Psychological Bulletin, 59*, 77-80.
- Q4 Hollinger, R. C., & Clark, J. P. (1983). *Theft by employees*. Lexington, MA: D.C. Heath.
- Hunt, S. T. (1996). Generic work behavior: An investigation into the dimensions of entry-level, Q4 hourly job performance. *Personnel Psychology, 49*, 51-83.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*, 1-19.
- Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgments of overall performance. *Journal of Applied Psychology, 86*, 984-996.
- Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7*, 238-257.
- Karren, R. J., & Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods, 5*, 337-388.
- Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations*. New York: Wiley.
- MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons' performance. *Organizational Behavior and Human Decision Processes, 50*, 123-150.
- MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1993). The impact of organizational citizenship behavior on evaluations of salesperson performance. *Journal of Marketing, 57*, 70-80.
- Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology, 79*, 475-480.
- Murphy, K. R., & Shiarella, A. H. (1997). Implications of the multidimensional nature of job performance for the validity of selection tests: Multivariate frameworks for studying test validity. *Personnel Psychology, 50*, 823-854.
- Naylor, J. C., & Wherry, R. J., Sr. (1965). The use of simulated stimuli and the JAN technique to capture and cluster the policies of raters. *Educational and Psychological Measurement, 25*, 969-986.
- O'Reilly, C. A., Chatman, J. A., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to person-organization fit. *Academy of Management Journal, 34*, 487-516.
- Q1 Organ, D. W. (1977a). A reappraisal and reinterpretation of the satisfaction-causes-performance hypothesis. *Academy of Management Review, 2*, 46-53.

- Organ, D. W. (1997b). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, *10*, 85-97.
- Q1 Pacolet, J., Van De Putte, I., Cattaert, G., & Coudron, V. (2002). *Plus est en vous herbekeken. Manpowerplanning in de zorgsector en de socioculturele sector: Deel 5. Synthesevoor de zorgsector in de Vlaamse Gemeenschap 1995-2020*, HIVA-K.U.Leuven, Leuven.
- Peterson, N. G., Mumford, M. D., Borman, W. C., Jeanneret, P. R., & Fleisman, E. A. (1999). *An occupational information system for the 21st century: The development of O*Net*. Washington, DC: American Psychological Association.
- Puffer, S. M. (1987). Prosocial behavior, noncompliant behavior, and work performance among commission sales people. *Journal of Applied Psychology*, *72*, 615-621.
- Q4 Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal*, *38*, 555-572.
- Q4 Rotundo, M. (2000). *The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach*. Unpublished Doctoral Dissertation, University of Minnesota.
- Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*, 66-80.
- Sackett, P. R., & DeVore, C. J. (2001). Counterproductive behaviors at work. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 145-164). Sage Publications.
- Q5 Schleicher, D. J., & Day, D. V. (1998). A cognitive evaluation of frame-of-reference rater training: Content and process issues. *Organizational Behavior and Human Decision-making Processes*, *73*, 76-101.
- Schmidt, F. L., & Kaplan, L. B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. *Personnel Psychology*, *24*, 419-434.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131-142.
- Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, *68*, 653-663.
- Van Scotter, J. R., & Motowidlo, S. J. (1996). Interpersonal facilitation and job dedication as separate facets of contextual performance. *Journal of Applied Psychology*, *81*, 525-531.
- Vey, M. A., & Campbell, J. P. (2004). In-role or extra-role organizational citizenship behavior: Which are we measuring? *Human Performance*, *17*, 119-135.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, *90*, 108-131.
- Werner, J. M. (2000). Implications of OCB and contextual performance for human resource management. *Human Resource Management Review*, *10*, 3-24.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189-205.

Received 10 July 2006; revised version received 3 January 2007

Author Queries

JOB NUMBER: 474

JOURNAL: JOOP

- Q1** We have changed the year for the references Organ, (1977) to Organ, (1977a) and Organ, (1977b) in list as per style. Please check and provide the respective reference citations in text part.
- Q2** We have made a change to this sentence. Please review our edit.
- Q3** Reference Katz and Kahn's (1979) has been cited in text but not provided in the list. Please supply reference details or delete the reference citation from the text.
- Q4** References Bennett and Robinson (2000), Cooper and Richardson (1986), George and Jones (1997), Hollinger and Clark (1983), Hunt (1996), Puffer (1987), Robinson and Bennett (1995), Rotundo (2000) are provided in the list but not cited in the text. Please supply citation details or delete the reference from the reference list.
- Q5** Please provide publisher location details for reference, Sackett and DeVore (2001).