Original Article

Disentangling Task and Contextual Performance

A Multitrait-Multimethod Approach

Evangelia Demerouti,[1] Despina Xanthopoulou,[2] Ioannis Tsaousis,[3] and Arnold B. Bakker[4]

[1]Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, The Netherlands, [2]School of Psychology, Aristotle University of Thessaloniki, Greece, [3]Department of Psychology, University of Crete, Rethymno, Greece, [4]Department of Work & Organizational Psychology, Erasmus University Rotterdam, The Netherlands

Abstract. This study among 244 employees and their colleagues working in various sectors investigated the dimensionality of self-ratings and peer-ratings of task and contextual performance, using the scales of Goodman and Svyantek (1999). By applying the multitrait-multimethod approach, we examined the degree to which responses to performance items are influenced by the trait (task or contextual performance) and the method factors (self- or peer-ratings). Results of confirmatory factor analyses showed that while the two performance dimensions (i.e., traits) can be differentiated, responses to their items are influenced by the method factor. Specifically, peer-ratings explain more variance in task performance, while self-ratings explain more variance in contextual performance. Moreover, the measurement of task and contextual performance is invariant across self- and peer-ratings. Finally, the positive relationships between task and contextual performance on the one hand and work-related flow on the other hand are of equal strength. These findings support the validity of the performance measure but also highlight some impact of the method factors.

Keywords: contextual performance, multitrait-multimethod approach, task performance, work-related flow

Capturing and predicting performance at work is no trivial pursuit. Task performance and contextual performance are generally perceived as important performance dimensions that are relevant for all kinds of jobs (Motowidlo, Borman, & Schmit, 1997). Because self-reports are subject to social desirability (Harris & Schaubroeck, 1988), researchers often use other-ratings of performance. However, other-ratings are also subject to biases (Cheung, 1999). One unclear issue is to what extent we need to differentiate between the information contained in self-ratings and other-ratings just like we differentiate between task and contextual performance. Our goal is to examine whether items capturing task and contextual performance form different (but related) dimensions, and whether self-ratings and peer-ratings of performance items exhibit substantial overlap.

The *first* objective of the present study is to investigate the degree to which responses to performance items are influenced by the trait (task or contextual performance) and the method factors (self- or peer-ratings). To meet this objective, simultaneous confirmatory factor analyses for items capturing both performance dimensions and both measurement ways will be conducted. We use the scale of Goodman and Svyantek (1999), which was proven to have satisfactory psychometric qualities (i.e., reliability, factorial validity). The confirmatory factor analysis will be used to test the hypothesis that the responses on performance items underlie the two performance dimensions of task and contextual performance, as well as the two measurement ways of self-ratings and peer-ratings. This is based on the criteria of Campbell and Fiske (1959) for multitrait-multimethod (MTMM) matrices.

The *second* objective of this study is to examine the pattern of relationships of the performance dimensions with work-related flow, a phenomenon that has started to attract increased research attention in the literature (Demerouti & Fullagar, 2013). We focus on flow at work as it represents the total involvement and immersion in one's task (Csikszentmihalyi, 1990) and has been shown to relate to both task and contextual performance (Bakker, 2008; Demerouti, 2006). Instead of examining the simple overlap between the constructs, as was done in earlier research, we will test their relationship in the MTMM matrices in order to uncover the measurement error-free relationships.

This study focuses on two explicit performance dimensions and raters that are commonly used in field organizational research. It adds to earlier MTMM studies (e.g., Conway, 1999) that focused either on *performance dimensions* captured by performance appraisal systems (namely job dedication, interpersonal facilitation, technical-administrative task performance, leadership task performance,

DOI: 10.1027/1866-5888/a000104

and overall performance) or on dimensions of managers' performance (namely administrative competence, human relations, and technical competence; Mount, Judge, Scullen, Sytsma, & Hezlett, 1998) by examining the impact of trait and method factors on the *responses to the items* rather than the dimensions of performance. Furthermore, an additional advantage of this study is that it examines the invariance of the items measuring task and contextual performance across raters. This is of particular importance because support for invariance suggests that different raters evaluate the same performance items in a similar way and thus, self- and peer-ratings of performance are comparable.

The Multitrait-Multimethod Model

According to Campbell and Fiske (1959), each validation process should be based on the following criteria: (1) both convergent and discriminant validity are required; (2) each scale that is used for measurement purposes is a trait-method unit (i.e., any particular trait is measured in a way that is not specific to that trait); and (3) in order to estimate the contribution of traits and methods it is imperative to test more than one trait and more than one method. These criteria are simultaneously satisfied when MTMM matrices are applied. In a full MTMM design, the analysis partitions the communality of the variables (i.e., the performance items) into variance caused by the trait and the method factors. Rather than examining the overlap among measures of the same construct for evidence of convergence, the analysis assesses the association of each measured performance item with the latent variable, and demonstrates significant convergent validation when the items obtain significant loadings on the appropriate trait factors (Floyd & Widaman, 1995). A significant factor loading on the relevant method factor indicates that the answer on the item is significantly influenced by method variance.

In the MTMM model, the method factors reflect the assumption that the method-specific influences are unidimensional within each method across traits. Every item has a value in the corresponding trait factor that contains no measurement error or method-specific influences. The trait variance is evidenced when the correlation between the methods (raters) assessing the same trait is high while evidence of method variance is present when the correlation between ratings of different traits made by the same method (rater) is high (Mount et al., 1998). The correlations between different traits are set free indicating discriminant validity, while the method factors are not correlated (Höfling, Schermelleh-Engel, & Moosbrugger, 2009).

Task Versus Contextual Performance

If we recall different types of jobs and the various tasks that are embedded in these jobs, we realize that it is not that easy to find an overall definition of performance that is applicable across situations. *Task* or *in-role* performance is relevant for every job and is defined as those officially required outcomes and behaviors that directly serve the goals of the organization (Motowidlo & Van Scotter, 1994). Among other things, task performance includes meeting company objectives and effective sales presentations (Behrman & Perreault, 1982). Note that the definition we provided emphasizes the instrumentality of performance for organizational goals. While this is certainly very important it does not describe the whole range of human performance at work.

Every employee also displays extra-role behaviors (Morrison, 1994). *Contextual* or *extra-role* performance or citizenship behavior is defined as discretionary behaviors on the part of an employee that are believed to directly promote the effective functioning of an organization without necessarily directly influencing the employee's productivity (MacKenzie, Podsakoff, & Fetter, 1991). Organ (1997) proposed that contextual performance is the best term for describing such activities. Contextual performance is defined as actions that go beyond what is stated in formal job descriptions and that increase organizational effectiveness highly depends on such extra-role actions (MacKenzie et al., 1991). One promising approach that has gained increased attention in the literature is to explicitly differentiate forms of citizenship according to the intended beneficiary. Examples of such conceptualizations are Williams and Anderson (1991), who suggested that citizenship behaviors can be directed toward the benefit of other individuals, or toward the benefit of the organization as a whole. Another example is Organ's (1988) taxonomy, who (originally) proposed that organizational citizenship behavior consists of altruism, courtesy, conscientiousness, civic virtue, and sportsmanship which are all parts of an overall helping dimension (Podsakoff, Whiting, Podsakoff, & Blume, 2009). A third example is that of Rupp and Cropanzano (2002), who suggested that employees might also direct their citizenship behavior toward their supervisor.

Several studies have shown that task and contextual performance can be differentiated empirically (e.g., Turnley, Bolino, Lester, & Bloodgood, 2003). In addition, Xanthopoulou, Bakker, Heuven, Demerouti, and Schaufeli (2008), who measured overall levels as well as episodes of task and contextual performance using the scale of Goodman and Svyantek (1999), further supported previous findings on the empirical distinction between these two traits. Like in the current study, task performance was measured with items that assessed employees' current performance (e.g., "Achieves the objectives of the job," "Plans and organizes to achieve objectives and meet deadlines") and promotion expectations (e.g., "Appears suitable for a higher level role," "Meets criteria for promotion"), while contextual performance was measured with the scale of altruism (e.g., "Helps others when their workload increases," "Takes initiative to orient new employees to the department even though not part of his/her job description), which represents citizenship behavior toward individuals. Hence, *Hypothesis 1*: After controlling for the method factors, responses to the items of task and contextual performance can be explained by different underlying (trait) dimensions of performance.

Self-Ratings Versus Peer-Ratings of Performance

Although self-reported measures give us some indication of individual job performance, they are flawed and hold only a tenuous to modest relationship with actual behavior and performance (Dunning, Heath, & Suls, 2004). Several studies have found that people find it difficult to analyze themselves objectively enough and to give accurate information about themselves (DeNisi & Shaw, 1977; Levine, Flory, & Ash, 1977). The human ability to self-report is subject to numerous factors that can influence the accuracy of the evaluation. Examples of this are intelligence (Freund & Kasten, 2012), high achievement status, and internal locus of control (Mabe & West, 1982), and the self-enhancement bias (Harris & Schaubroeck, 1988; Mabe & West, 1982; Sackeim, 1983; Taylor & Brown, 1988). Empirical evidence from meta-analytic studies suggests that the correlations between self-ratings and peer-ratings are modest and range between .19 (Conway & Huffcutt, 1997) and .36 (Harris & Schaubroeck, 1988). Thornton (1980) concluded that individuals have a significantly different view of their own job performance than that held by other people. However, Vandenberg, Lance, and Taylor (2004) showed that other-reports of contextual performance also include systematic biases and concluded that self-reports of contextual performance were most appropriate operationalizations. Thus, both self- and other-ratings of performance seem to contain valuable information.

Harris and Schaubroeck (1988) provide three possible explanations regarding why low correlations may emerge between self- and other-ratings of performance. First, because of egocentric biases or defensiveness, individuals inflate their ratings leading to a restricted range and lower correlations. Moreover, according to attribution theory, individuals are inclined to attribute good performance to personal factors and poor performance to environmental factors. The second explanation refers to the fact that raters coming from different organizational levels measure performance differently and weight the diverse performance dimensions in a different way. Therefore, they can disagree on their overall ratings but may agree on specific aspects of performance. Finally, Harris and Schaubroek suggested that peers have more chances than supervisors to observe the employee's behavior. These three explanations are not mutually exclusive but can operate simultaneously. Defensiveness may be the cause for the differences between self- and peer-ratings, whereas observational opportunities may be the cause for differences between peer- and supervisor-ratings.

Conway and Huffcutt (1997) concluded that the different sources (individual, peers, supervisors) have somewhat different perspectives on performance and that agreement within sources (e.g., peer and peer) is higher than agreement between sources (e.g., peer and self). Moreover, self- and other-ratings contain specific information that cannot be captured by objective measures. For instance, unlike other raters and objective data, self-raters are aware of the moves they undertake for career preservation and enhancement as well as their behavior toward different people at work (cf. Hall, 1990). As one reviewer keenly noted – since contextual performance targets different entities, these entities are naturally best placed to report on these behaviors. Therefore, we expect that:

Hypothesis 2: Next to the trait factors, responses to performance items will be explained by the method factors of self- and peer-ratings.

Performance and Flow at Work

The term flow was first defined by Csikszentmihalyi (1977, p. 36) as "the holistic sensation that people feel when they act with total involvement." In the ensuing definitions of flow, Csikszentmihalyi has mentioned additional aspects including the balance between challenges and skills, increased likelihood to learn new skills, and being in control of own actions (Csikszentmihalyi, 1990). Drawing on the flow literature, Bakker (2008) recognized three states as the core elements of the flow experience at work: absorption, enjoyment, and intrinsic motivation. The first element, *absorption*, refers to total concentration and immersion in the activity (Csikszentmihalyi, 1990; Ghani & Deshpande, 1994). Employees in flow are so focused on what they do that they forget everything else around them. The second core element of flow is *enjoyment* (Csikszentmihalyi, 1990; Trevino & Webster, 1992). Employees, who enjoy their work and feel happy, make a very positive judgment about the quality of their working life (cf. Veenhoven, 1996). This enjoyment or happiness is the outcome of cognitive and affective evaluations of the flow experience (cf. Diener, 2000). The third and final element, *intrinsic motivation*, refers to the state in which people do what they do "even at great cost, for the sheer sake of doing it" (Csikszentmihalyi, 1990, p. 3; Trevino & Webster, 1992).

According to Engeser and Rheinberg (2008), flow should be associated with better task but also contextual performance for two reasons. First, flow is a highly functional state (balance between skills and challenges, and being in control) that should in itself foster performance. Second, individuals experiencing flow are more motivated to carry out further activities, and in order to experience flow again, they will recreate challenging tasks for themselves (Ceja & Navarro, 2012). Thus, flow could be seen as a motivating force for excellence. Additionally, when in flow, people are very concentrated, and they invest all available cognitive and energy resources. According to the performance episodes model (Beal, Weiss, Barros, & MacDermid, 2005), resource allocation to the task is crucial for successful performance. If employees cannot allocate all of their energetic and motivational resources to the task at hand, for example because they are constantly interrupted by colleagues, e-mails, or phone calls, they cannot perform optimally. Thus, allocating, replenishing, and conserving self-regulatory resources is critical for successful performance during performance episodes (Beal et al., 2005).

Several studies have documented the relationship between flow and performance (Nakamura & Csikszentmihalyi, 2005). For example, Demerouti (2006) found that flow predicted both task and contextual performance among conscientious employees. Similarly, Bakker (2008) found that flow was positively related to task and contextual performance. However, these studies looked at the simple correlations between flow and peer ratings of performance without correcting for measurement error and without a partitioning of the variance of performance into trait and method factors. On the basis of the literature, our final hypothesis is:

> *Hypothesis 3*: After controlling for the method factors, task and contextual performance will relate positively to (self-ratings of) flow at work.

Method

Participants and Procedure

The sample, materials, and the collection of data are described in more detail in Demerouti (2006) and Bakker, Demerouti, and Verbeke (2004). As both studies used similar methodologies to collect data from heterogeneous working populations, we pooled the data for the analyses. Briefly, the participants in the present study were employed in various sectors and job positions. After their initial agreement, supervisors from several small- and large-sized companies in the Netherlands received a letter explaining the goal of the study. Several days later, they were contacted by phone to ask how many questionnaires could be sent to their organization. In total, 454 questionnaires were distributed to 21 different companies. In the accompanying letter, the anonymity of participants' responses was emphasized. In addition, each participant was instructed to ask one colleague to rate his/her task and contextual performance. Participants and their colleagues could return the questionnaires separately with stamped envelopes. The dyads were matched by using a unique code for each set of questionnaires.

A total of 259 employees and their colleagues filled out and returned the questionnaire. Due to the fact that there were missing data from 15 colleagues, these 15 dyads were excluded resulting in a total of 244 dyads of employees and their colleagues (54% response rate). The final sample of employees included 119 males (49%) and 125 females (51%). Participants' age ranged from 20 to 62 years with an average of 38 years (*SD* = 10.60). The majority of the sample had a university degree (29%) or higher vocational training (32%). Organizational tenure was on average 5 years (*SD* = 8.1), and 69% of the sample was full-time employed. Since we did not ask participants at which company or institute they worked, we do not know how many different organizations were finally included in the study. However, what we do know is that most participants worked with people (70%); 25% worked primarily with information and 5% worked primarily with "things."

Measures

Task performance was assessed with the nine-item scale developed by Goodman and Svyantek (1999). An example item is: "I demonstrate expertise in all job-related tasks." The same items were used by participants' colleagues to rate participants' performance but adapted accordingly (e.g., "Demonstrates expertise in all job-related tasks"). Participants and their colleagues were asked to indicate the extent to which they found each statement characteristic of themselves or the participant respectively, on a scale ranging from (1) not at all characteristic to (7) totally characteristic.

Contextual performance was measured with the seven-item scale reported in Goodman and Svyantek (1999). These authors based their instrument on Smith, Organ, and Near's (1983) organizational citizenship behavior measure, and named their scale "altruism" (i.e., citizenship behavior toward individuals). Example item is: "Help/s other employees with their work when they have been absent." Participants and their colleagues were again asked to rate their own or participants' contextual performance respectively, using the same answering categories as for the task performance scale.

Work-Related Flow was measured with Bakker's (2008) work-related flow inventory. This scale consists of 13 items that measure the three underlying dimensions of the work-related flow construct: *absorption* ("When I am working, I forget everything else around me," four items), *work enjoyment* (e.g., "I do my work with a lot of enjoyment" four items), and *intrinsic motivation* (e.g., "I get my motivation from the work itself, and not from the reward for it," five items). Participants were asked to rate their own flow experiences on a scale ranging from (0) never to (6) always.

Analysis Strategy

We performed simultaneous confirmatory factor analyses (CFA) for both rating methods to test the hypothesis that the task and contextual performance items underlie the two distinct dimensions. Analyses were performed with the AMOS 20.0 software package (Arbuckle, 2011) using maximum likelihood estimation. Following (Marsh & Hocevar, 1988; see also Marsh & Grayson, 1995), we fit the following alternative models to the data. First, we tested a one-factor model (M1), where all items measuring task and contextual performance – as rated by both employees themselves and their colleagues – loaded on a single "performance" latent factor. Second, we tested a two-factor, trait model (M2), which hypothesizes that the variation in the performance items is explained by the underlying traits (i.e., the two performance dimensions) plus errors.

Thus, this model does not recognize any differentiation between measurement methods (i.e., type of rating). This model included two correlated latent factors (one for task and one for contextual performance). The task performance latent factor was operationalized by the task performance items (included as observed variables) as rated both by participants and their colleagues, and the contextual performance latent factor was operationalized by the contextual performance items as rated both by participants and their colleagues. Third, we tested a two-factor, method model (M3), which assumes that the factor structure of performance is determined by the method of measurement (i.e., self- and peer-ratings) and not by the underlying performance dimensions. This model included two correlated latent factors: one representing the self-ratings (which was operationalized by the task and contextual performance items that were rated by the participants), and one representing the peer-ratings (that was operationalized by the task and contextual performance items that were rated by participants' colleagues).

Next, we tested a four-factor model (M4), which included four latent factors: one representing the task performance items as rated by employees, one representing the task performance items as rated by their peers, one representing the contextual performance items as rated by employees, and one representing the contextual performance items as rated by their peers. All four latent factors were allowed to correlate with each other. Finally, we fit the MTMM model (M5) to the data. The MTMM model included all performance items and two categories of latent factors: (1) the two performance facets (traits) that were correlated, and (2) the two ratings of performance (methods) that were not correlated. Each item (manifest indicator) had two loadings: one on the trait factor and one on the method factor. High covariance between self- and peer-ratings of performance suggests high convergence validity. We did not correlate the method factors because the correlation between methods was already modeled in the trait factors. Further, it is important to note that in all examined models we included correlated uniquenesses among matching items (i.e., we allowed the errors of each item that was rated by different raters to correlated freely). As suggested by Marsh and Hocevar (1988; Marsh et al., 2011), uniqueness refers to the part that is unique to each item that is evaluated by different raters, and includes measurement error that is shared by the different raters. Controlling for uniquenesses prevents from resulting in inflated correlations between constructs measured with the same items. Finally, we examined the invariance of the MTMM model across raters, as described in the Results section.

The second objective of the present study was to investigate the construct validity of the MTMM model. To this respect, we added flow (as a latent variable operationalized by absorption, work enjoyment, and intrinsic motivation as manifest variables) in the MTMM model, as well as covariances between the latent flow factor on the one hand, and task and contextual performance latent factors on the other hand. We were particularly interested in examining the strength of the relationships between flow at work on the one hand and task and contextual performance (while controlling for the method of rating) on the other hand.

The fit of the models to the data was assessed with the chi-square (χ^2) statistic, the Standardized Root Means Square Residuals (SRMR) and the Root Mean Square Error of Approximation (RMSEA). In addition, two, less sensitive to sample size, fit indices were used: the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). For each of these statistics, values of .90 are acceptable and of .95 or higher are indicative of good fit (Hu & Bentler, 1999). RSMEA below .05 in combination with SRMR values below .09 indicate excellent fit, whereas values below .08 and .10, respectively, indicate good fit (Byrne, 2001). Alternative models were compared on the basis on the chi-square difference test.

Results

Before proceeding with the main analyses for testing the study hypotheses, we ran preliminary CFAs to test the factor structure of task and contextual performance separately for self- and peer-ratings. More specifically, we compared the fit of the proposed two-factor model (i.e., which included one task performance latent factor with the respective task performance items loading on this factor, and one contextual performance latent factor with the respective contextual performance items loading on this factor) to the fit of an alternative one-factor model (i.e., where all task and contextual performance items loaded on a general performance factors) and a modified two-factor model (i.e., where some changes were implied based on the modification indices). These results are presented in Table 1.

Table 1 shows that the proposed two-factor model fits the data significantly better than the alternative one-factor model both for self-ratings, $\Delta\chi^2(1) = 150.48$, $p < .001$, and for colleague ratings, $\Delta\chi^2(1) = 226.99$, $p < .001$. However, despite the superiority of the two-factor model for both methods, some fit indices did not satisfy the criteria for good fit. For this reason, we examined the modification indices for potential cross-loadings. As concerns the self-ratings of performance, modifications indices clearly suggested that the error of item 5 of the task performance scale (i.e., "Could manage more responsibility than typically assigned") related both to the task performance (M.I. = 4.33) and the contextual performance (M.I. = 6.95) latent factors. Furthermore, modification indices suggested that the errors of items 6 ("Makes innovative suggestions to improve the overall quality of the department"; M.I. = 7.28) and 7 ("Willingly attends functions not required by the organization, but helps in its overall image"; M.I. = 5.42) of the contextual performance scale could relate to the task performance latent factor. As concerns the colleague-ratings, inspection of the modification indices allowed similar conclusions with regard to these items. In addition, it was shown that item 6 of the contextual performance scale loaded on the task performance latent factor. All in all, these findings suggest that these items are problematic and could be excluded from

Table 1. Results of confirmatory factor analyses for self-ratings and colleague-ratings ($N_{employees} = 244$; $N_{colleagues} = 244$)

Model	χ^2	*df*	RMSEA	CFI	TLI	SRMR	AIC
Self-ratings							
M1: one-factor	463.53	104	.12	.84	.81	.067	527.527
M2: two-factor	313.05	103	.09	.91	.89	.060	379.054
M3: two-factor modified	144.20	64	.07	.95	.95	.043	190.201
Colleague-ratings							
M4: one-factor model	496.39	104	.13	.85	.83	.067	560.392
M5: two-factor model	269.40	103	.08	.94	.93	.057	335.399
M6: two-factor modified	63.28	64	.00	1.00	1.00	.025	117.281

Notes. df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; AIC = Akaike information criterion.

Table 2. Descriptives, Cronbach's α, and correlations between the study variables ($N = 244$)

	Variables	Mean	*SD*	α	1	2	3	4	5	6	7
	Self-ratings										
1	Task performance	5.12	1.02	.90	–						
2	Contextual performance	4.89	1.16	.86	.68**	–					
3	Absorption	3.93	1.10	.79	.32**	.31**	–				
4	Work enjoyment	5.02	1.02	.88	.29**	.32**	.53**	–			
5	Intrinsic motivation	4.12	1.06	.77	.24**	.29**	.64**	.71**	–		
6	Flow total	4.36	0.92	.90	.33**	.36**	.84**	.85**	.90**	–	
	Peer-ratings										
7	Task performance	5.47	1.02	.92	.42**	.31**	.13*	.19**	.21**	.20**	–
8	Contextual performance	5.32	1.16	.90	.24**	.41**	.11	.12	.18**	.16*	.68**

Note. $^*p < .05$; $^{**}p < .05$.

further analyses as we needed to use clear traits in order to examine the impact of trait and method factors to the item responses. To confirm whether a two-factor model without these problematic items fit better to the data, we tested this alternative model for self-ratings (M3) and colleague-ratings (M6). As Table 1 shows the two-factor alternative model shows the best fit to the data (both M3 and M6 resulted in the lowest AIC value as compared to the other models). On the basis of these analyses, we decided to use this alternative measurement solution in the remaining of the paper.

Descriptive Statistics

Table 2 presents descriptive statistics, reliabilities, and correlations between the study variables. All scales showed good reliabilities with Cronbach's alphas values varying between .77 and .90. Self-ratings of task performance were significantly correlated with peer-ratings of task performance ($r = .42$, $p < .01$). Similarly, the correlation between self- and peer-ratings of contextual performance was positive and significant ($r = .41$, $p < .01$). Also, Table 2 shows that self-ratings of task and contextual performance correlate equally high as peer-ratings of the same dimensions (i.e. $r = .68$, $p < .01$). Finally, the total flow score correlated significantly, yet moderately with self- and peer-ratings of both task and contextual performance.

Validity Analyses

Table 3 presents the fit indices of the competing models for the MTMM analysis. The results of the chi-square difference test indicated that the MTMM model (shown in Figure 1) was superior to all alternative models apart from the four-factor model. Results concerning the four-factor model (that represents four latent factors, each distinguishing self- and peer-ratings of task and contextual performance) suggested that the four latent factors interrelated significantly with one another. Specifically, self-ratings of task and contextual performance were high inter-related ($\varphi = .76$, $p < .001$), just like peer-ratings of task and contextual performance ($\varphi = .74$, $p < .001$). Furthermore, self- and peer-ratings of task performance ($\varphi = .44$, $p < .001$), as well as self- and peer-ratings of contextual performance ($\varphi = .42$, $p < .001$) correlated significantly. As expected, the other correlations were substantially lower (self-ratings of task performance and peer-ratings of contextual performance, $\varphi = .21$, $p < .001$; self-ratings of contextual performance and peer-ratings of task performance $\varphi = .33$, $p < .001$) but still significant.

Table 3. Fit indices of alternative factor structures of task and contextual performance (N = 244)

Model	χ^2	*df*	RMSEA	CFI	TLI	SRMR	Comparison	$\Delta\chi^2$	Δdf
M1: one-factor model	2,078.37	286	.161	.565	.506	.158	–	–	–
M2: two-factor, trait model	1,761.45	285	.146	.642	.592	.151	M1-M2	316.92***	1
M3: two-factor, method model	812.59	285	.087	.872	.854	.068	M1-M3	1,265.78***	1
M4: Four-factor model	401.80	280	.042	.970	.966	.043	M2-M4	1,359.65***	5
							M3-M4	410.79***	5
M5: MTMM model	393.60	259	.046	.967	.959	.080	M4-M5	8.20	21
M6: MTMM and flow	510.50	335	.046	.961	.952	.080	–	–	–
M7: MTMM and flow constrained	510.69	336	.046	.961	.953	.080	M6-M7	0.19	1
								ΔCFI	ΔRMSEA
M8: MTMM model – equal factor loadings	434.92	272	.050	.960	.953	.132	M5-M8	.007	.004
M9: MTMM model – equal intercepts	467.66	285	.051	.956	.949	.135	M8-M9	.004	.001
M10: MTMM model – equal uniquenesses	543.10	298	.058	.941	.935	.133	M9-M10	.015	.007

Notes. MTMM = multitrait-multimethod; RMSEA = root mean square error of approximation; CFI = confirmatory fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean square residual; *p < .05; **p < .01; ***p < .001.

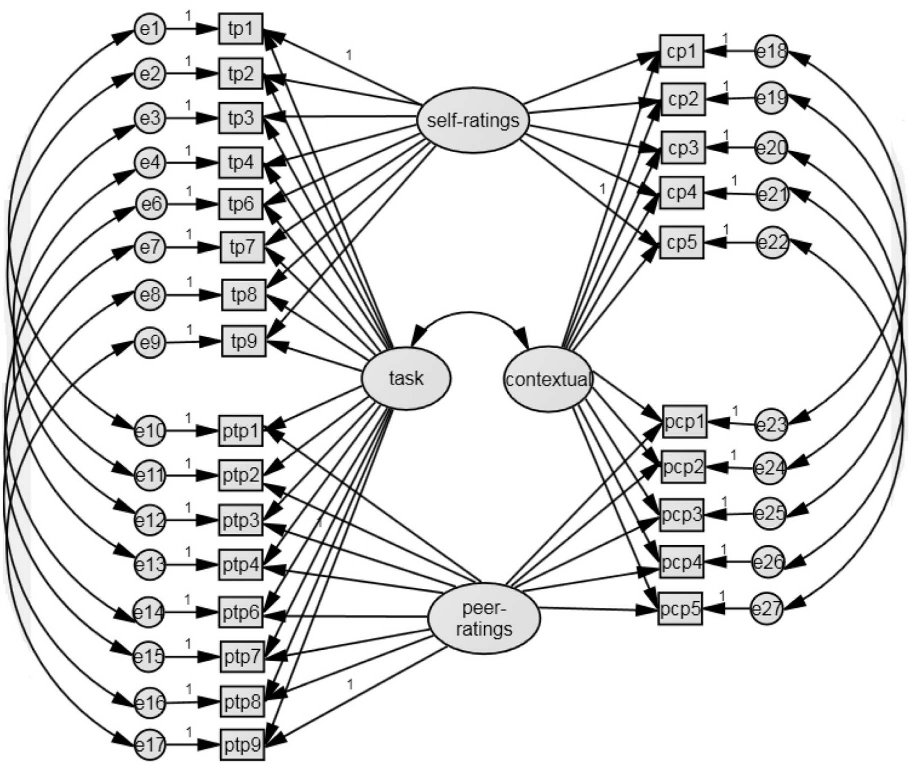

Figure 1. The multitrait-multimethod model (tp = task performance; cp = contextual performance; ptp = peer-rating of task performance; pcp = peer-rating of contextual performance).

Table 3 shows that the MTMM model fit equally well to the data as the four-factor model, $\Delta\chi^2(21) = 8.20$, *ns*. This suggests that both trait- and method-related differentiation is substantial when measuring job performance. Furthermore, the estimated correlation between the trait factors was $\varphi = .38$ ($p = .04$). This correlation accounts for the common variance of the two trait factors (after correcting for the methods) and it demonstrates discriminant validity because it is moderate (cf. Höfling et al., 2009). Note that the estimated correlation between the traits in the MTMM model is lower than their correlations in the four-factor model. Further, we calculated the communality (R-square) of the item loadings in the MTMM model, in order to estimate the variance explained by each factor. The model explains a total of 29%–80% (M = 57.4%) of the variance in task performance items and a total of 44%–93% (M = 71.9%) of the variance in contextual performance items. For self-rated task performance items, the trait factor explains an average of 30% of the variance in items ratings, versus 47.23% for the method factor. For peer-rated task performance items the trait factor explains an average of 12.79% for the trait factor versus 49.28% for the method factor. Similarly, for self-rated contextual performance items the trait factor explains an average of 9.6% of the variance and 48% of the variance for the method factor. Finally, for peer-rated contextual performance items the trait factor explains an average of 35.2% of the variance and 26.6% of the variance for the method factor. All in all, these results provide support for Hypothesis 1 and 2 as we could clearly differentiate between task and contextual performance as well as between self-ratings and peer-ratings of performance.

Test of Invariance Across Raters

To further test the measurement invariance of the MTMM model across raters, we tested a series of alternatives to the main MTMM model (M6). As suggested by Vandenberg and Lance (2000), we tested three alternative models (see Table 3): (1) a model identical to the MTMM model, where we constrained the loading of matching items to be equal across raters (M8; Equal Factor Loadings); (2) a model identical to M8, where we also constrained item intercepts to be equal across raters (M9; Equal Intercepts); and (3) a model identical to M9, where we also constrained uniquenesses to be equal across raters (M10; Equal Uniquenesses). For evaluating invariance across these models, we examined changes in CFI (ΔCFI). According to Cheung and Rensvold (2002), a CFI decrease of .01 or higher indicates a significant decrement in model fit and lack of invariance across raters. The simulation results of Chen (2007) confirm the adequacy of the .01 criterion and suggest that changes in RMSEA should also be examined with a cut-off score of .015. As shown in Table 3, all alternative models fit equally well to the data, while only M10 (where next to factor loadings and intercepts, also uniquenesses were constrained to be equal across raters) was found to exceed the ΔCFI threshold. Even though these results generally support the global invariance of item's uniquenesses, the change in the CFI value led us to further probe for potential differences across raters. This examination indicated that the measurement errors tend to be slightly higher for self-reports than for peer reports. Nevertheless, the invariance of the MTMM model significantly holds on the basis of the insignificant changes in the RMSEA of the examined models.

Job Performance and Work-Related Flow

As depicted in Table 3, the MTMM model in which flow was also included (M6) showed acceptable values for all fit indices. For this model, McDonald's (1999) omega reliability coefficients were estimated for both, task performance (0.83) and contextual performance (0.74) scales. Furthermore, both covariances between task performance and flow ($\varphi = .29$, $p < .01$), and between contextual performance and flow ($\varphi = .21$, $p < .01$) were significant and positive. These results provide support for Hypothesis 3. In a final step we compared the strength of these relationships, by comparing M6 to a model identical to M6 where equality constraints were added to the covariances between task performance and flow, and contextual performance and flow (M7; Table 3). The chi-square difference test showed that both models fit the data equally well, $\Delta\chi^2(1) = .19$, *ns*, suggesting that the relationships of task and contextual performance with flow are of equal strength.

Discussion

The aim of this study was to examine the degree to which responses to items that capture performance are influenced by the specific trait of performance that is assessed and by the method that is used to measure it. In order to investigate this, we used two dimensions of the performance scale reported in Goodman and Svyantek (1999), namely the task and contextual performance dimensions. These dimensions are relevant for every job and frequently investigated in organizational research. Information about job performance was provided by the individual self as well as a direct colleague. In addition, we examined the relationships of the self-ratings and peer-ratings of both performance dimensions to flow at work. To examine these goals we applied the MTMM approach.

Taken together, we found some encouraging results regarding the differentiation between the two performance dimensions. Although the bivariate correlation between task and contextual performance was high, the estimated correlation in which the method factors were also controlled for was rather low which supports discriminant validity. Similarly, the measurement of task and contextual performance was found to be invariant across raters, that is employees and their colleagues. The validity of the performance dimensions was further supported by the fact that their relationship to work-related flow does not vary significantly. However, we also found some critical results regarding the role of the method factor. Specifically, there is more method variance in the peer-ratings task performance and in the self-ratings of contextual performance.

Theoretical Contributions

Our MTMM findings showed that the model differentiating between the two performance (i.e., task and contextual) dimensions was better than the model that refrained from doing this. In line with earlier research (e.g., Motowidlo & Van Scotter, 1994), this finding indicates that irrespective of who rates task and contextual performance, they represent different sub-dimensions of performance (even at the item-level of analysis). However, their estimated correlation (i.e., after correcting for measurement error and controlling for the rater) was moderate ($\varphi = .38$), their bivariate correlation was high ($r = .68$), and they related identically with flow at work. The good news of these results is that task and contextual performance show discriminant validity. The critical note is that they contain a substantial amount of overlapping information. Although they are still far from being considered as redundant dimensions of performance, researchers should be aware of their overlap. Viswesvaran and Ones (2000) showed in their review that different performance dimensions of individual job performance are positively correlated and that this positive manifold suggests the presence of a general factor, which represents a common variance shared across all the dimensions.

Furthermore, analysis showed that the measurement of task and contextual performance is invariant across raters meaning that (1) both employees and their colleagues perceived the items of the scales in the same way (equal factor loadings); (2) that individuals with the same score on the latent construct would obtain the same score on the observed variable regardless of who is the rater (equal intercepts); and (3) that the relationships between the items are similar across raters. These findings underscore the validity of the task and contextual performance as measured by the scale of Goodman and Svyantek (1999) as different raters seem to interpret and rate the items in the same way.

MTMM analysis further showed some interesting results regarding the role of the method factor in the responses to performance items. First, the model differentiating between self-ratings and peer-ratings of performance (irrespectively of the dimension) was better than the model that considered these as identical methods. This means that self-ratings contain information that is different from peer-ratings which agrees with earlier research. For instance, Conway (1999) analyzed multitrait-multimethod (rater) performance appraisal matrices and concluded that rating system characteristics showed low-to-moderate levels of correlation. In total the MTMM model explained almost 58% of the variance in the task performance items and substantially more, namely 72%, of the variance in the items of contextual performance.

Thus, it seems that we can better explain the factors that influence responses to contextual performance ratings than to task performance ratings. Moreover, we found that for some dimensions the method factor explained substantial variance in the item responses. Specifically, peer-ratings explained more variance in task performance, while self-ratings explained more variance in contextual performance. This means that although both raters understand the performance items in the same way, responses to specific performance dimensions are influenced more by a certain rating source than by another. According to Mount et al. (1998), one potential explanation for this finding is halo, which occurs when the rater's ratings are heavily influenced by an overall evaluation of the rate. Another potential cause of strong method effects is that raters from different levels observe different aspects of performance and may also use different standards when judging performance.

A final contribution of the present study is that it shows a relationship between self-ratings of work-related flow on the one hand and task and contextual performance (using both self- and peer-ratings) on the other hand. As hypothesized, our MTMM analysis showed that both dimensions of performance related positively to flow irrespective of the rater. This finding signifies the importance of full engrossment in an activity for both types of performance (Beal et al., 2005; Engeser & Rheinberg, 2008). Moreover, both performance dimensions showed the same strength of relationship with flow, which is a sign of convergence rather than discriminant validity.

The critical question that arises from these findings is what researchers could do in future research to overcome the problem that self- and other ratings of (task and contextual) performance contain different information. A first solution would be to move toward more specific and perhaps more observable dimensions of performance, since raters seem to agree more when they rate specific rather than global dimensions of performance/aspects of the job. Similarly, Furnham and Stringfield (1998) in their study among task-oriented teams whose members' performance was rated by the self, the superior, team-peers, and a consultant who was part of the team found that congruence among raters was higher for specific, observable behaviors, and lower for less observable cognitive aspects of performance. Although the utilization of global performance dimensions like task and contextual performance leads to shorter questionnaires, one should consider the danger of using less reliable and valid operationalizations. A second solution would be to use supervisor rather than peer ratings of (particularly task) performance as supervisors seem to be better trained to evaluate the performance of employees, and thus more reliable and less biased raters (e.g., Arvey & Murphy, 1998). The utilization of supervisor-ratings would not reduce the discrepancy between self-ratings and supervisor-ratings. Rather it would result in more reliable data. A final solution would be to simply accept the limitations of the measurement of these performance dimensions and not only to include in studies both self-ratings and peer-ratings of performance but also to report the findings of both self-reports and peer-ratings. In this way, readers can see whether and how results differ when data is collected with self-reports and with other-ratings.

Limitations and Future Research

The first limitation of this study is its reliance on cross-sectional data. Although it provides a useful consideration of the factor structure of the different performance dimensions, while taking into account different rating methods, it cannot address the predictive validity of flow. This would require the separation in time of the measurement of flow and the performance dimensions.

A second limitation concerns the choice to collect data from individual employees and their peers as these provide biased information. To overcome this problem, one should aim for the additional collection of more objective information, like objective absenteeism (Demerouti, Bouwman, & Sanz-Vergel, 2011), financial returns of the organization (Xanthopoulou, Bakker, Demerouti, & Schaufeli, 2009), or objective sales performance (Bakker, Van Emmerik, & Van Riet, 2008; Verbeke, Belschak, Bakker, & Dietz, 2008). Moreover, it would be valuable to combine self- and peer-ratings of performance with ratings by supervisors, subordinates, and customers, who have access to other type of information which sometimes is not available either to the employees themselves or their peers.

Future research could investigate daily fluctuations in (peer and self-ratings of) task performance and contextual performance. Diary methods help reduce retrospective bias and can help achieve more objective assessments, since with these methods we capture life as it is lived (Bolger, Davis, & Rafaeli, 2003). As less measurement errors will influence the answers of individuals about their own or their colleague's behavior on a specific day, it will be more likely to find higher overlap between self- and other-ratings.

Finally, future research needs to examine whether these findings generalize also to other performance measures. Because some items had no clear loading on one of the two traits and were therefore eliminated, it is possible that the results cannot be generalized. Replication of the findings using different scales and raters would strengthen our conclusions.

Conclusion

By applying MTMM analysis, we showed that while task and contextual performance can be differentiated, responses to their items are influenced by the method factor. However, there is substantial method variance for the peer-ratings of task performance and for the self-ratings of contextual performance, which needs caution. Furthermore, the measurement of task and contextual performance is invariant across employees and their colleagues, while the two performance dimensions show similar positive relationships to work-related flow. Because of these findings we propose that researchers should measure both self- and other-ratings of performance dimensions and model these using MTMM models in which both the traits and the methods are differentiated. In this way, an accurate prediction of performance can be achieved, which is advantageous for both research and practice.

References

Arbuckle, J. (2011). *AMOS 20.0 user's guide*. Crawfordville, FL: AMOS Development Corporation.

Arvey, R. D., & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141–168.

Bakker, A. B. (2008). The work-related flow inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior, 72*, 400–414.

Bakker, A. B., Demerouti, E., & Verbeke, W. (2004). Using the Job Demands – Resources model to predict burnout and performance. *Human Resource Management, 43*, 83–104.

Bakker, A. B., Van Emmerik, H., & Van Riet, P. (2008). How job demands, resources, and burnout predict objective performance: A constructive replication. *Anxiety, Stress, and Coping, 21*, 309–324.

Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology, 90*, 1054–1068.

Behrman, D. N., & Perreault, W. D., Jr. (1982). Measuring the performance of industrial salespersons. *Journal of Business Research, 70*, 355–370.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.

Byrne, B. M. (2001). *Structural equation modelling with Amos. Basic concepts application and programming*. Mahwah, NJ: Erlbaum.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Ceja, L., & Navarro, J. (2012). "Suddenly I get into the zone": Examining discontinuities and nonlinear changes in flow experiences at work. *Human Relation, 65*, 1101–1127.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.

Cheung, G. W. (1999). Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology, 52*, 1–36.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness of fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233–255.

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer and self-ratings. *Human Performance, 10*, 331–360.

Csikszentmihalyi, M. (1977). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.

Conway, J. M. (1999). Distinguishing contextual performance from task performance for managerial jobs. *Journal of Applied Psychology, 84*, 3–13.

Demerouti, E. (2006). Job characteristics, flow and performance: The moderating role of conscientiousness. *Journal of Occupational Health Psychology, 11*, 266–280.

Demerouti, E., Bouwman, K., & Sanz-Vergel, A. I. (2011). Job resources buffer the impact of work-family conflict on absenteeism in female employees. *Journal of Personnel Psychology, 10*, 166–176.

Demerouti, E., & Fullagar, C. J. (2013). Experiencing flow in the workplace and what individuals and organizations can do to foster it. In R. J. Burke & C. L. Cooper (Eds.), *The fulfilling workplace: The organization's role in achieving individual and organizational health* (pp. 71–89). Farnham, UK: Gower.

DeNisi, A. S., & Shaw, J. B. (1977). Investigation of the uses of self-reports of abilities. *Journal of Applied Psychology, 62*, 641–644.

Diener, E. (2000). Subjective well-being, the science of happiness and a proposal for a national index. *American Psychologist, 55*, 34–43.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*, 158–172.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286–299.

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*, 296–321.

Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360 Feedback examining self, manager, peers, and consultant ratings. *Human Relations, 4*, 517–530.

Ghani, J. A., & Deshpande, S. P. (1994). Task characteristics and the experience of optimal flow in human-computer interaction. *Journal of Psychology, 128*, 381–391.

Goodman, S. A., & Svyantek, D. J. (1999). Person-organization fit and contextual performance: Do share values matter. *Journal of Vocational Behavior, 55*, 254–275.

Hall, D. T. (1990). Career development theory in organizations. In D. Brown & L. Brooks (Eds.), *Career choice and development* (pp. 422–454). San Francisco, CA: Jossey-Bass.
Harris, M. M., & Schaubroeck, J. (1988). A meta analysis of self-supervisor, self-peer and peer-supervisor ratings. *Personnel Psychology, 41*, 43–62.
Höfling, V., Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three methods. *Methodology, 5*, 99–111.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equational Modeling, 6*, 1–55.
Levine, E. L., Flory, A. P., & Ash, R. A. (1977). Self assessment in personnel selection. *Journal of Applied Psychology, 62*, 428–435.
Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280–296.
MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons. *Organizational Behavior and Human Decision Processes, 50*, 123–150.
Marsh, H. W., & Grayson, D. (1995). Latent-variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues and applications* (pp. 177–198). Newbury, CA: Sage.
Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73*, 107–117.
Marsh, H. W., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G., & Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology, 103*, 701–732.
McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
Morrison, E. W. (1994). Role definitions and organizational citizenship behavior: The importance of the employee's perspective. *Academy of Management Journal, 37*, 1543–1567.
Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance, 10*, 71–83.
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.
Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater, and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557–576.
Nakamura, J., & Csikszentmihalyi, M. (2005). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). Oxford, UK: Oxford University Press.
Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome*. Lexington, MA: Lexington Books.
Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance, 10*, 85–97.
Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Blume, B. D. (2009). Individual-and organizational-level consequences of organizational citizenship behaviors: A meta-analysis. *Journal of Applied Psychology, 94*, 122–141.
Rupp, D. E., & Cropanzano, R. (2002). Multifoci justice and social exchange relationships. *Organizational Behavior and Human Decision Processes, 89*, 925–946.
Sackeim, H. A. (1983). Self-deception, self-esteem, and depression: The adaptive value of lying to oneself. In J. Masling (Ed.), *Empirical studies of psychoanalytic theories* (pp. 101–157). Hillsdale, NJ: Erlbaum.
Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology, 68*, 653–663.
Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*, 193–210.
Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33*, 263–270.
Trevino, L. K., & Webster, J. (1992). Flow in computer-mediated communication. *Communication Research, 19*, 539–573.
Turnley, W. H., Bolino, M. C., Lester, S. W., & Bloodgood, J. M. (2003). The impact of psychological contract fulfillment on the performance of in-role and organizational citizenship behaviors. *Journal of Management, 29*, 187–206.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of measurement of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.
Vandenberg, R. J., Lance, C. E., & Taylor, S. C. (2004). A latent variable approach to rating source equivalence: Who should provide ratings on organizational citizenship behavior dimensions? In D. L. Turnipseed (Ed.), *A handbook of organizational citizenship behavior: A review of "good soldier" activity in organizations* (pp. 105–138). New York, NY: Nova Science.
Veenhoven, R. (1996). Happy life-expectancy. *Social Indicators Research, 39*, 1–58.
Verbeke, W. J., Belschak, F. D., Bakker, A. B., & Dietz, B. (2008). When intelligence is (dys)functional for achieving sales performance. *Journal of Marketing, 72*, 44–57.
Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment, 8*, 216–226.
Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management, 17*, 601–617.
Xanthopoulou, D., Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2009). Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology, 82*, 183–200.
Xanthopoulou, D., Bakker, A. B., Heuven, E., Demerouti, E., & Schaufeli, W. B. (2008). Working in the sky: A diary study on work engagement among flight attendants. *Journal of Occupational Health Psychology, 13*, 345–356.

Evangelia Demerouti

Department of Industrial Engineering & Innovation Sciences
Eindhoven University of Technology
Human Performance Management Group
P.O. Box 513
5600 MB Eindhoven
The Netherlands
Tel. +31 40 247-5669
Fax +31 40 243-7161
E-mail e.demerouti@tue.nl