ICLTIBM 2017
7th International Conference on Leadership, Technology, Innovation and Business Management

EVALUATION OF TASK AND CONTEXTUAL PERFORMANCE: A MULTITRAIT-MULTIMETHOD APPROACH

Safiye Şahin (a)*, Uğur Yozgat (b)
*Corresponding author

(a) Kırklareli University, Kırklareli, Turkey, E-mail: safiye.sahin@medeniyet.edu.tr
(b) Marmara University, İstanbul, Turkey

This study aims to investigate the discriminant and convergent validity of task and contextual performance, using the scales of Goodman and Svyantek (1999) with self-ratings and supervisor-ratings. The total sample included 486 employees and their supervisors, working in the public hospitals. Firstly, we ran preliminary CFAs to test the factor structure of task and contextual performance separately for self-ratings and supervisor-ratings. Then, we examined the convergent and discriminant validity of task and contextual performance by applying the multitrait-multimethod approach. Results of multitrait-multimethod analyses indicated that the two performance dimensions (task and contextual performance) can be differentiated and the measurement of task and contextual performance is invariant across self-ratings and supervisor-ratings. According to the results of MTMM analyses two hypotheses are supported. Task and contextual performance can be differentiated and the measurement of task and contextual performance is invariant across self-ratings and supervisor-ratings. This research highlights some aspects of performance appraisal.

© 2019 Published by Future Academy www.FutureAcademy.org.UK

Keywords: Contextual performance, task performance, multitrait-multimethod approach.

1. Introduction

As is known, performance appraisal is one of the most critical responsibilities of human resources management. As a result of the prevalence of self-managing work teams and special work teams, evaluation is not only the supervisor's duty any more. Rather, in order to evaluate individual performance, organizations are increasingly benefitting from multiple sources. Today performance appraisal does not only reflect the supervisor's perception but it also reflects peers and customers' perception (Miller and Cardy, 2000). In 360-degree performance appraisal system, information is gathered from supervisors, peers, subordinates, customers and employees themselves (Milliman et al., 1994). Thus, this way of evaluation provides not only a top-down evaluation, but also it provides both the same level and bottom-up evaluation opportunity. One of the most important outcomes of this method is that it prevents subjective evaluation. Moreover, it shows how the employee perceives his own performance and helps him to see how he is perceived by those around him (Miliman et al., 1994). On the other hand, 360-degree performance appraisal systems draw various criticisms. The most important of those is applying the person himself as a source of information.

2. Literature Review and Hypotheses

2.1. Task and Contextual Performance

According to Borman and Motowidlo (1997), there are two types of employee job performance behavior. These are role performance and extra-role performance. At the same time, role performance and extra-role performance are also called task performance and contextual performance respectively. Task performance expresses duties and responsibilities of a job which makes it different from others (Jawahar & Carr, 2006). Being related to expertise and mechanics of the job (Borman, 2004), task performance focuses on basic technical details in the job and means artifice of the task required to carry out a job successfully (Van Scotter & Motowidlo, 1996). Despite being mostly focused on task performance to reach organizational goals, in the long run researchers have realized that various activities contributing to actualization of organizational goals exist. Borman and Motowidlo (1993) called these activities which support implementation of the organizational activities successfully contextual performance.

2.2. Self-Ratings and Manager-Ratings of Performance

How and by whom the performance evaluation is made is intensely discussed in the literature. According to a variety of studies, people find it hard to analyze themselves objectively enough and to give accurate information about themselves (DeNisi & Shaw, 1977; Levine, Flory, & Ash, 1977). People tend to overvalue their own performances (Thornton, 1980). Some studies on this effect have revealed that this tendency has different aspects in western and eastern cultures. In eastern cultures, individuals are observed to rate their own performances lower than their superiors and this was called "humility effect". In western cultures, the results show just the opposite (Farh, Dobbins, & BarSchiuan, 1991). In contrast to varying results in this study and this type of studies, Farh and Webel (1986) have observed that self-enhancement bias decrease in individual's self-evaluation once he is informed that an objective performance criterion is used for comparison. Within the scope of this study, self-enhancement bias is believed to decrease since

individuals know that information will be gathered through another sources as well and there will be a comparison. Other than that, studies in which self-evaluations are compared with different sources provide more accurate information about the validity of self-evaluation. When compared to performance in the past and peer-ratings and various psychological tests, self-rating is stated to be as much valid as these methods (Shrauger & Osberg, 1981). On the other hand, DeNisi and Shaw (1977), in their research with 144 university students, evaluated students' skills via two methods (self-evaluation and skills tests). They have found out there is a low correlation between the results of self-evaluation and skills tests, and reported that self-evaluation method cannot replace skills tests. People's ability to evaluate themselves, is subject to a great variety of factors which can affect the accuracy of evaluation such as intelligence, high success status, internal locus of control and self-enhancement bias. According to Thornton (1980), individuals have quite different opinions about their own job performances than people around them have. Conway and Huffcutt (1997) have concluded that different sources (individual, peer, supervisor) have quite different perspectives of performance and agreement within sources (peer and peer) is more notable than the agreement between sources (peer and self). In studies on task oriented teams whose members' performances are evaluated by themselves, their supervisors, peers and a consultant who is a part of the team, Furnham and Stringfield (1998), on the other hand, have found out rater agreement is higher in specific and observable behaviors and lower in cognitive dimensions of performance. They have argued that evaluation of performance by a supervisor would be more effective than by a peer. The reason for this is supervisors are better educated to evaluate employees; from this point of view they are more reliable and less biased raters.

2.3. The Multitrait-Multimethod Model

In social sciences, there is known to be a number of measurement tools which measure a trait. In theory, the same trait which is measured by different methods is supposed to be correlated with one another. One of the best ways to carry out valid measurements for any observed variant is to use different methods to discover different traits. For this purpose, Campbell and Fiske (1959) have put forth multitrait-multimethod model (MTMM). This approach includes use of multiple methods and traits. Both convergent validity and discriminant validity studies are made in structure validity studies with MTMM and these studies show effect of applied methods (Campell & Fiske,1959). Convergent validity is put into perspective when there is one trait and two methods while discriminant validity is put into perspective with multiple traits and a single method (Höfling et al., 2009). Different structures are defined with the same methods, afterwards measurements for each structure are obtained with each method and their correlation with each other is calculated. Final correlation coefficients are defined as being one of either convergent validity coefficient or discriminant validity coefficient. Correlation between the measurements of the same structure is observed by using different methods in convergent validity coefficient and correlation is expected to be high. On the other hand, in discriminant validity coefficient, correlation between different structures is observed by using the same measurement method or different methods; correlation is expected to be lower than convergent validity coefficient (Campell & Fiske,1959).

3. Methodology

3.1. Research Goal

Objective of this study is to analyze contextual and task performance evaluations by self-ratings and supervisor ratings via multitrait-multimethod analyses.

3.2. Research Hypotheses

H1: After controlling for the method factors, trait dimensions of performance which are named as the task and contextual performance can be differentiated.

H2: The measurement of task and contextual performance will be invariant across self-ratings and supervisor-ratings.

3.3. Sample and Data Collection

486 medical personnel and 61 supervisors who work for public hospitals in İstanbul constitute the sample of the study. Questionnaires were handed over and collected by the surveyor. Medical staff and staff in charge of the services in the hospitals were interviewed in person, were informed about the study, and how to fill the questionnaire was explained to them elaborately. Afterwards names and questionnaire numbers of every volunteer who accepted to participate in the study from each service were noted down on a small notepaper and mentioned paper was handed over to the superior. The supervisor wrote down the questionnaire number of whichever subordinate he would evaluate on the questionnaire form and destroyed this paper. By this way, data of the supervisor and subordinate were combined. In order to maintain confidentiality of the information, sealed envelopes were handed over to both the supervisor and the subordinates, and participants submitted the questionnaire form to the surveyor in a closed envelope. SPSS 23 and AMOS 23 were used for analysis of the data.

3.4. Measures

The scale which was developed to measure job performance by Goodman and Svyantek (1999) was used in the study. While subordinates evaluated their own performance through this scale, supervisors evaluated their subordinates' performances through the same scale. Originally first 16 points of this scale constitute contextual performance while the last 9 constitute task performance. Contextual performance constitutes of two sub dimensions. However, only "altruism" sub dimension comprising of 7 points was used as part of this study. Here are a couple of points from the scale: "Helps other employees with their work when they have been absent." and "Takes initiative to orient new employees to the department even though it is not a part of his/her job description." On the other hand, task performance does not have any sub dimensions. "Achieves the objectives of the job." and "Performs well in the overall job by carrying out tasks as expected." are examples of the points in this scale. Likert Scale, which is originally a 7-point scale, was transformed into a 6-point scale as part of this study (1= Strongly Disagree; 6= Strongly Agree). Lower points from the scale show lower performance level while higher points show higher performance level.

3.5. Results

Firstly, we ran preliminary CFAs to test the factor structure of task and contextual performance separately for self-ratings and supervisor-ratings. We examined six models. These are:

- One-factor model where all task and contextual performance items loaded on a general performance factors for self-ratings and supervisor-ratings (Model 1 and Model 4).

- Two-factor model 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 for self-ratings and supervisor-ratings (Model 2 and Model 5).

- A modified two-factor model where some changes were implied based on the modification indices for self-ratings and peer-ratings (Model 3 and Model 6).

These results are presented in Table 1.

Table 01. Results of CFA for Self-ratings and Supervisor-ratings (N=487)

Model	X²/df	RMSEA	CFI	TLI	RMR	AIC
Self-ratings						
M1: One-factor	6,727	0,109	0,857	0,835	0,094	763,637
M2: Two-Factor	5,554	0,097	0,888	0,869	0,093	638,050
M3: Two-factor modified	2,913	0,063	0,978	0,969	0,037	111,738
Supervisor-ratings						
M4: One-factor	11,079	0,144	0,866	0,845	0,074	1216,245
M5: Two-Factor	8,215	0,122	0,905	0,889	0,80	912,165
M6: Two-factor modified	2,705	0,059	0,989	0,985	0,016	108,323

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 1 presents that the proposed two-factor model fits the data better than the one-factor model both for self-ratings and for supervisor-ratings. However, some fit indices did not satisfy the criteria for good fit for the two-factor model for both methods. Therefore, we examined the modification indices for potential cross-loadings. Then, 2th, 4th, 6th, 7th, 13th, 14th and 16th items were excluded from analyses and these two factor modified models (M3 and M6) were used in the remaining of the paper.

After testing the factor structure of task and contextual performance separately for self-ratings and supervisor-ratings, we examined descriptive statistics, reliabilities, and correlations between the research variables.

Table 02. Descriptives, Chronbach's α, and Correlations between Variables

Variables	Mean	SD	α	1	2	3	4
Self-ratings							
1. Contextual Performance	4,85	,89	.761	1			
2. Task Performance	4,88	,84	.879	.643**	1		
Supervisor-ratings							
3. Contextual Performance	4,84	,82	.887	**.173****	.220**	1	
4. Task Performance	4,79	,93	,952	.137**	**.206****	.808**	1

Note. **p<0,01

Table 2 shows that descriptive statistics, reliabilities, and correlations between the research variables. All scales showed good reliabilities with Cronbach's alphas values varying between 0,76 and 0,95. Self-ratings of contextual performance were significantly correlated with supervisor-ratings of contextual performance (r = .173, p< .01). Similarly, the correlation between self-ratings and supervisor-ratings of task performance was positive and significant (r = .220, p < .01). Also, Table 2 shows that supervisor-ratings of contextual and task performance correlate higher (r = .808, p < .01) than self-ratings of the same dimensions (r = .643, p< .01).

Secondly, we examined five models. These are:

1. M1: All items loaded on a single performance latent factor.

2. M2: This model included two correlated latent factors which were named as task and contextual performance. There was no differentiation between two measurement methods (self and supervisor ratings) for this model.

3. M3: This model included two correlated latent factors which were named as self-ratings and supervisor-ratings performance. There was no differentiation between two performance dimensions (task and contextual performance) for this model.

4. M4: This model included four correlated latent factors. Task performance items as rated by employees loaded on a self-ratings task performance latent factor, contextual performance items as rated by employees loaded on a self-ratings contextual performance latent factor, task performance items as rated by supervisors loaded on a supervisor-ratings task performance latent factor and contextual performance items as rated by supervisors loaded on a supervisor-ratings contextual performance latent factor.

5. M5: MTMM Model. This model is shown in Figure 1.

Table 03. Fit indices of alternative actor srtuctures of contextual and task performance (N=486)

Model	X²	df	RMSEA	CFI	TLI	RMR	Comparison	ΔX²	Δdf
M1: one-factor model	2600,009	136	0,193	0,620	0,572	0,233	-	-	-
M2: two-factor, trait model	2336,627	134	0,184	0,660	0,612	0,231	M1-M2	263,38***	2
M3: two-factor, method model	658,266	134	0,090	0,919	0,908	0,055	M1-M3	1941,74***	2
M4: four-factor model	236,392	129	0,041	0,983	0,980	0,036	M2-M4 M3-M4	2100,23*** 421,87***	5 5
M5: MTMM model	195,728	107	0,041	0,986	0,980	0,052	M4-M5	40,664**	22

Table 3 shows that four-factor model and the MTMM model fit well to the data, ΔX² (22) = 40,664, p<0,01. Moreover, the estimated correlation between the trait factors in the MTMM model was ₵ = .42 (p = .015). According to Höfling et al., 2009; if the estimated correlation between the traits in the MTMM model is moderate to low, it will demonstrate discriminant validity. Besides, the estimated correlation between the traits in the MTMM model (₵ = .42) is lower than their correlations in the four-factor model (₵ = .73 for self-ratings and ₵ = .83 for supervisor-ratings). So, these findings support the discriminant validity in our research.
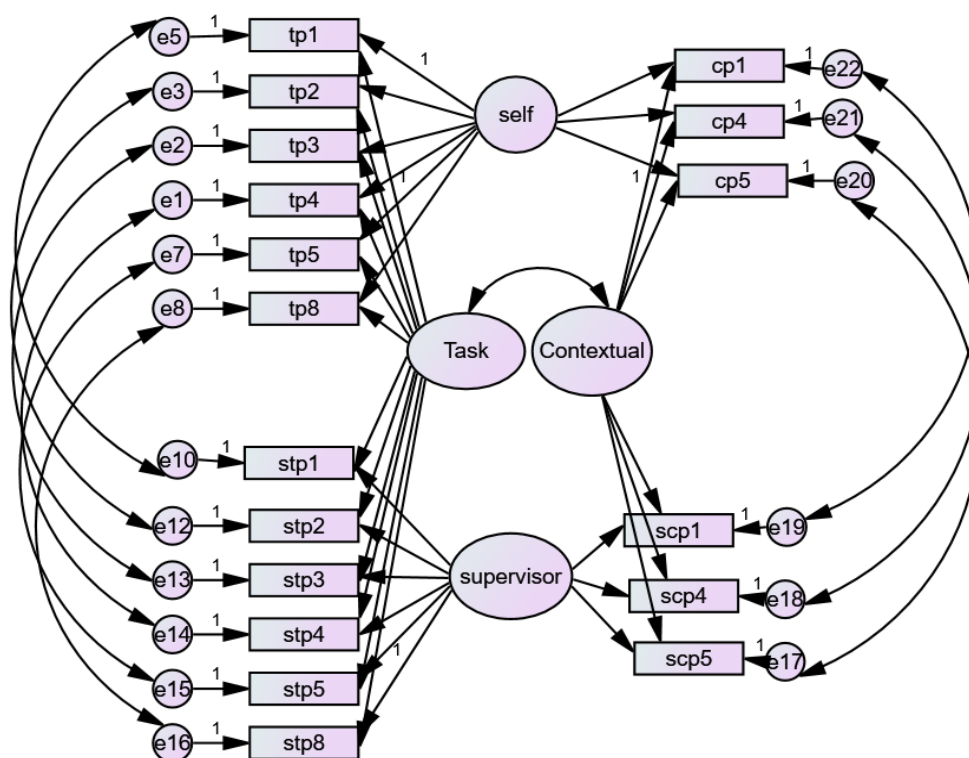
Figure 01. The Multitrait-Multimethod Model
Note: tp =task performance; **cp** = contextual performance; **stp** = supervisor-rating of task performance; **scp** =
supervisor-rating of contextual performance

In order to estimate the variance explained by each factor, we calculated the communality (R-square) of the item loadings in the MTMM model:

- The model explains a total of 14%–70% (M = 38%) of the variance in **task performance** items and a total of 12%–54% (M = 35%) of the variance in **contextual performance** items.

- For **self-rated contextual performance** items the trait factor explains an average of 17% of the variance and 71% of the variance for the method factor.

- For **self-rated task performance** items, the trait factor explains an average of 56% of the variance in items ratings, versus 50% for the method factor.

- For **supervisor-rated contextual performance** items the trait factor explains an average of 53% of the variance and 67% of the variance for the method factor.

- For **supervisor-rated task performance** items the trait factor explains an average of 19% for the trait factor versus 86% for the method factor.

Table 04. The Communality of the item loadings in the MTMM model

	Our Research	
	Trait Factor	**Method Factor**
Self-rated Task Performance	**%56**	**%50**
Supervisor-rated Task Performance	%19	**%86**
Self-rated Contextual Performance	%17	**%71**
Supervisor-rated Contextual Performance	**%53**	**%67**

The measurement of task and contextual performance was found to be invariant across raters, that is employees and their supervisors. However, there is more method variance in the supervisor-ratings task performance and in the self-ratings of contextual performance (Table 4).

4. Conclusion

Performance appraisal is one of the most critique issue for organizations. It is known that organizations' success depends on their employees' performance. However, the measurement of employee's performance has some complexities. One of them is about the types of employee's job performance behavior. In the literature, there are two performance behaviors, task and contextual performance. The other one is about that how and by whom the performance appraisal is made. In 360-degree performance appraisal system, information is gathered from supervisors, peers, subordinates, customers and employees themselves. In order to disambiguate the performance appraisal, we examined contextual and task performance evaluations by self-ratings and supervisor ratings via multitrait-multimethod analyses. We tested two hypotheses about discriminant and convergence validity of job performance scale by Goodman and Svyantek (1999). Results of MTMM analyses provide support for two hypotheses. Task and contextual performance can be differentiated and the measurement of task and contextual performance is invariant across self-ratings and supervisor-ratings. This research highlights some aspects of performance appraisal.

References

Borman, W. C. (2004). The concept of organizational citizenship. *Current Directions in Psychological Science, 13*(6), 238-240.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. N. Schmitt, & W. C. Borman (Eds.), *Personnel Selection* (p. 71-98.), Josey-Bass, San Francisco.

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

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.

Demerouti, E., Xanthopoulou, D., Tsaousis, I., & Bakker, A.B. (2014). Disentangling task and contextual performance : A multitrait-multimethod approach. *Journal of Personnel Psychology, 13*(2), 59-69.

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

Farh, Jiing-Lih L., & Werbel, J. D. (1986). Effects of purpose of the appraisal and expectation of validation on self-appraisal leniency. *Journal of Applied Psychology, 71,* 527-529.

Farh, Jiing-Lih L., Dobbins, G. H., & Bar-Shiuan, C. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. Workers. *Personnel Psychology, 44,* 129-147.

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 jobperformance ratings: A study of 360 Feedback examining self, manager, peers, and consultant ratings. *Human Relations, 4,* 517–530.

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

Höfling, V., Schermelleh-Engel, K., & Moosbrugger, H. (2009). Analyzing multitrait-multimethod data: A comparison of three methods. *Methodology, 5,* 99-111.

Jawahar, I., M., & Carr, D. (2007). Conscientiousness and contextual performance. *Journal of Managerial Psychology, 22*(4), 330-349.

Levine, E. L., Flory, A. P., & Ash, R. A. (1977). Self assessment in personnel selection. *Journal of Applied Psychology, 62,* 428–435.

Miller, J. S., & Cardy, R. L. (2000). Self-monitoring and performance appraisal: rating outcomes in project teams. *Journal of Organizational Behavior, 21,* 609-626.

Milliman, J. F., Zawacki, R. A., Norman, C., Powell, L., & Kirksey, J. (1994). Companies evaluate employees from all perspectives. *Personnel Journal, 73,* 99-103.

Shrauger, S. J., & Osberg, T. M. (1981). The relative accuracy of selfprediction and judgements by others in psychological assessment. *Psychological Bulletin, 90*, 322-351.

Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33,* 263–270.

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.