Short Version of Self-Assessment Scale of Job Performance

Érika Guimarães Soares de Azevedo Andrade[1], Fabiana Queiroga[2], and Felipe Valentini[3]

1 Universidade Salgado de Oliveira (Universo) (Brasil)
2 Centro Universitário de Brasília (UniCEUB) (Brasil)
3 Universidade São Francisco (USF) (Brasil)

Título: Versión reducida de la Escala de Autoevaluación del Desempeño en el Trabajo.
Resumen: Este artículo tiene como objetivo reducir la Escala de autoevaluación de Desempeño en el trabajo, como también controlar el direccionamiento de respuesta y aprobación, utilizando la técnica de viñetas y ítems invertidos. Se utilizó el banco de datos de la escala original, compuesta por 20 ítems divididos en dos factores: tarea y contexto. Para la reducción, fueron elegidos los 10 ítems con mayores cargas factoriales y *thresholds*. La escala reducida fue estimada por un factor general y dos dimensiones específicas: Tarea y Contexto, representando un modelo bifactor, con indicadores de ajustes adecuados (RMSEA = 0,05; TLI = 0,98). Para controlar el direccionamiento de respuesta y aprobación, fue realizada una colecta de datos, en la cual las respuestas fueron recodificadas y realizados análisis factoriales con la finalidad de realizar una comparación de los resultados con y sin la utilización de viñetas y ítems invertidos. Los resultados apuntaron a que las viñetas mejoraron las cargas factoriales de los análisis, y que los ítems invertidos no tuvieron mejores resultados además de las viñetas.
Palabras clave: Desempeño en el trabajo; Autoevaluación de rendimiento; Direccionamiento de respuesta.

Abstract: This paper aims to reduce the job performance self-assessment scale as well as control the response bias and acquiescence ~~bias~~ using vignettes anchors and inverted items. The original scale database was composed of 20 items divided into two factors: task and context. For the reduction, the ten items with higher factor loads and thresholds were chosen. The reduced scale was estimated by a general factor and two specific dimensions: task and context, representing a bifactor model, with adequate adjustment indicators (RMSEA = .05; TLI = .98). To control response bias and acquiescence, a second study was carried out, in which the responses were recoded and factor analyses were performed in order to make a comparison of the results with and without the use of the vignettes and inverted items. The results indicated that the vignettes improved the factorial loads; however, the reversed items did not perform better than the vignettes.
Keywords: Job performance; Self-evaluation of performance; Response bias.

Introduction

Job performance is an essential variable for organizational psychology. It is characterized as a dynamic process, which receives constant influence from the environment, the individual themself, and the workgroup. Without individual performance, there is no team performance, no unit performance, no organizational performance, no economic sector performance. Despite its importance, define what is individual performance is not something easy (Campbell & Wiernik, 2015). During the 1990s, multidimensional models of performance as were discussed by Borman and Motowidlo (1997) and Campbell (1990). From these sources, a consensus developed that individual job performance should be defined as things that people do, actions the individual performed to achieve a desired goal or target within the organization (Campbell, 2012; Campbell & Wiernik, 2015).

Therefore, to understand what performance is, a conceptual distinction needs to be established between individual goal-directed behaviors and the results of these behaviors. The latter is more quantifiable and relates to the products delivered or attained at work. Performance behaviors relate to the actions performed by the individual to produce a result that meets an organizational goal sought. In this sense, performance, from the viewpoint of organizational behavior, is an individual characteristic and does not necessarily pre-

sent a perfect correlation with the results at work (Campbell, 2012). An example that characterizes this conceptual distinction is that of the car salesman, who has an excellent result in relation to the number of sales when this sale takes place under favorable economic conditions (such as the reduction of the Tax on Manufactured Products). However, the same seller can maintain the same performance, that is, present good behaviors related to the sale of cars and obtain inferior results (sell fewer cars) during the month when the tax returns to the standard rate.

More up-to-date performance models are presented in recent literature, such as Campbell's (1990/2012) and Campbell and Wiernik (2015), which include eight characteristics, and the Koopmans et al. (2014), which include a counterproductive performance dimension in addition to task and context. To analyze job performance with a focus on individual behavior, the model that supports this present study is that of Sonnentag and Frese (2002), which makes a distinction between task-oriented performance and context-oriented performance. Used in models with other constructs, the Sonnentag and Frese (2002) model presents excellent prediction indicators with other variables (as found in Brazil by Paula and Queiroga (2015) with job satisfaction and organizational climate and Brandão et al. (2012), with the individual variables).

The task performance is related to the technical core of the organization, that is, to the production stage and how the activities of individuals collaborate on the technical issues of the company. Contextual performance, on the other hand, refers to work activities that do not directly contribute to the technical aspects of production but are embedded in the

broader levels of the social, organizational, and psychological environment. Moreover, contextual performance involves proactive and strategic behaviors. In short, task performance can be represented by the skills the individual learns to perform a task or develop a product. Contextual performance, then, is closer to the idea of organizational behavior and citizenship, in which commitment to the organization supports the provision of ideas and suggestions to improve work procedures to achieve the desired goals (Sonnentag & Frese, 2002).

Both individual (such as task proficiency, motivation to work, job satisfaction, job engagement) and job context-related variables (such as organizational climate, perceived support to work, leadership style) can predict performance (Obeidat Shatha et al., 2016). That is, the variables relate differently to the types of performance. Data from a meta-analysis by Bing et al. (2011) reinforce, for example, that political skills are better predictors of context-oriented than task-oriented performance.

Other individuals variables are job satisfaction and engagement. These constructs are similarly related to both task-oriented and context-oriented performance (Bowling et al., 2015; Edwards et al., 2008). But when one looks at the specific facets, the satisfaction with the type of job is more strongly related to task-oriented performance. In contrast, satisfaction with the supervisor is more related to context-oriented performance. A similar result was verified by Paula and Queiroga (2015) in a study that considered job satisfaction and organizational climate. Those authors identified that the predictive value of satisfaction with the type of job and support from the head (organizational climate variable) is higher for context-oriented performance. On the other side, engagement is a construct related to context-oriented performance (Bowling et al., 2015).

In summary, the studies show that more individual characteristics (such as cognitive skills, knowledge, length of experience, and personality traits) are more associated with task-oriented performance. In contrast, aspects more related to the work environment (such as organizational citizenship, job engagement, and organizational climate) would be variables more associated with context-oriented performance-oriented. Environmental variables exert a significant influence on performance because they impact the behavior of the individual at work and also on the individual variables (Huang & Su, 2016).

Illustrating these relationships, Coelho Junior and Borges-Andrade (2011), based on a multilevel model, studied the impact on the performance of individual variables (such as education, gender, and job satisfaction) and learning support in an indirect public management company. The authors found that the shared variance between first and second-level variables indicates that contextual factors, when analyzed jointly with individual factors, can explain significant performance variance related to the results the individual achieved.

Based on the task and context performance model (Sonnentag & Frese, 2002), a General Self-Assessment Scale of Work Performance was developed (Queiroga, 2009). Despite the broad acceptance of the model (Campbell & Wiernik, 2015), until that moment, there was no instrument for measuring individual performance that considered it. Thus, an instrument with 20 self-reported items was development, being eleven context-related (items that evoke the individual's proactivity and strategic action form), and nine task-related (which are items related to the execution of the tasks, based on the work techniques work).

Being an instrument with good psychometric indicators, in this study, we aim to propose a reduction of this scale from 20 to 10 items, so that it can be applied faster, with greater agility, decreasing the response time of the participant without losing the psychometric characteristics of the original scale. Thus, we expect to reduce the original scale from 20 to 10 items (distributed between task and context), maintaining satisfactory adjustment indicators. Shorter instrument versions are particularly crucial in a survey in work psychology, due to the high number of constructs often investigated. Therefore, reducing a scale in 50% yields the possibility to add another construct in future surveys.

Moreover, as is well known, the responses given to this type of self-reported scale are influenced by the subject's response style, which can lead to bias in the research results that use them (Primi et al., 2016). One threat is the acquiescence, also known as the "yea-saying" effect, which concept is related to the inclination to endorse positive categories of a Likert scale despite the item content. For instance, the items "I usually work hard" and "I am a lazy worker" are semantic antonyms (positive and negative worded), and, using a Likert scale, we expect opposite answers. However, a highly acquiescent subject will tend to agree with both inconsistently. Such a phenomenon can compromise the internal structure of the scores, adding an artificial general factor (Danner et al., 2015; Maydeu-Olivares & Coffman, 2006; Rammstedt & Farmer, 2013). In this sense, we suspect the response styles are biasing the internal structure of the Scale of Job Performance. Thus, the present study also aims to control response bias and acquiescence utilizing vignettes anchors and inverted items. Thus, the hypothesis postulates that the factor loadings will be higher for the models with control for the response style through the anchoring vignettes (hypothesis 1) and with control for acquiescence through inverted items (hypothesis 2).

This research is divided into two studies. Study 1 presents the reduction of the General Self-Assessment Scale of Job Performance. Study 2 aims to confirm the factorial structure of the instrument with and without control for response bias.

Study 1

Method

Participants

The first stage of this research included the use of the psychometric validation database of the original scale performed by Queiroga et al. (2015), composed of 1,617 participants, being 57.5% of banking employees and 42.5% employees from a joint-stock company in the oil sector. In both organizations, most respondents were male (80%) and had finished higher education (Queiroga et al., 2015).

Instruments

The Self-Assessment Scale of Job Performance (SJoP) was initially constructed with 20 items answered on a five-point frequency scale ranging from 1 (never) to 5 (always). This number of items was designed to evaluate several nuances of the dimensions of the Sonnentag and Frese (2002) model, and its structure presents items related to task-oriented and context-oriented performance. The task performance is aimed at the technical core of the company, and it is related to the skills learned and the behaviors expected to perform a specific job. An example of an item is: "I perform difficult tasks properly." Context performance is focused on the social and psychological support needed to achieve organizational goals. This dimension also involves proactive and strategic behaviors. An example of an item is: "I take initiatives to improve my results at work." The authors developed the original scale aiming the comparison between several organizations, which is especially useful in a practical context. Moreover, self-reports places performance assessments at the same level as other individual variables often used to explain performance. Thus, it is possible to analyze explanatory models when the assumptions of the multilevel analysis are not reached (Kozlowski & Klein, 2000). Furthermore, although self-reports are relatively biased because they tend to overestimate performance ratings, organizational assessments are also not devoid of contamination and deficiency (Edwards et al., 2008). The scale functionality has been tested in the original study of the Queiroga (2009), and the factor analyses indicated that the two dimensions explained 39.4% of the item variance. Besides, Cronbach's alpha coefficients were equal to 0.88 and 0.82 for the context and task factors, respectively.

Procedures

The five best items of each dimension were selected from the original scale database (Queiroga, 2009), that is, items with higher factor loads and better indicators of psychometric adjustments, totaling ten items with higher factor loadings. The structure was modeled with one general and two specific dimensions (one related to the context and one related to the task).

The parameters of the items were estimated through structural equation modeling. Based on the recommendations of Byrne (2013) and Hu and Bentler (1998), the following fit indices were analyzed: chi-square (tests the probability of the theoretical model adjusting to the data, and the higher the χ^2, the worse the goodness of fit); Root Mean Square Error of Approximation (RMSEA - should be inferior to .05, accepting coefficients as low as .08); Tucker-Lewis Index (TLI); Comparative fit index (CFI). CFI and TLI coefficients superior to 0.95 were considered acceptable. We also use Parsimony CFI (PCFI) for comparing models. The scores reliabilities were estimated through Composite Reliability and Hierarqical Omega (Primi et al., 2013; Valentini & Damásio, 2016), which are more robustious for measures with not homogeneous factor loadings. The analyses were performed through the software Mplus.

Results

The reduction of the number of items in the scale was performed based on factor loadings and thresholds. Thus, we tried to maintain the items with the highest factor loadings and various thresholds, to preserve items appropriate to the different levels of the psychological construct. After choosing the ten items, different models were tested based on the scale theory: single-factor model, two-factor model (with and without correlation), and also the bifactor model.

The single factor model was tested as it would be plausible to have an overall dimension that encompasses all the performance behaviors at work from the theoretical point of view. Considering the two-factor model (with and without correlation), the tests were performed as items cover behaviors related to the task performed and the context. And concerning the bifactor model, this was also tested for covering the two previous models, as it consists of a general factor and two specific dimensions.

Table 1. Goodness of fit indicators of a single factor, two factor (with and without correlation) and bifactor models of the Self-Assessment Scale of Job Performance.

Models	χ^2 (df)	TLI	CFI	PCFI	RMSEA (CI)
1. Two factors (with correlation)	808.29 (34)	.937	.952	.719	.118 (.111 − .125)
2. Two factors (no correlation)	8225.96 (35)	.493	.348	.384	.377 (.370 − .384)
3. Single factor (unidimensional)	1235.21 (35)	.904	.926	.720	.144 (.137 − .151)
4. Bifactor	201.30 (28)	.983	.989	.615	.061 (.053 − .069)

Although the correlated two-factor model presented a reasonable fit, the correlation between the two dimensions was very high (0.83 - value not shown in the table). This correlation is higher than the average of the factor loadings (0.75 for task and 0.78 for context). These results indicate that, in a first-order model, the scores of factors do not present evidence of discriminant validity. On the other hand, it can be verified that the uncorrelated two-dimensional model is not plausible, as the goodness of fit was not satisfactory.

Table 2. Non-standardized factor loadings of items of Self-Assessment Scale of Job Performance (short version).

Items	Description	Factor loadings (Standard error)		
		General Factor	Task	Context
Item 1	I perform hard tasks properly.	.59 (.02)	n.s.	
Item 2	I try to update my technical knowledge to do my job.	.67 (.02)		.09 (.04)
Item 3	I do my job according to what the organization expects from me.	.69 (.02)	.06 (.03)	
Item 4	I plan the execution of my job by defining actions, deadlines and priorities.	.67 (.02)	.67 (.14)	
Item 5	I plan actions according to my tasks and organizational routines.	.69 (.02)	.52 (.11)	
Item 6	I take initiatives to improve my results at work.	.82 (.01)		n.s.
Item 7	I seek new solutions for problems that may come up in my job.	.75 (.02)		.32 (.04)
Item 8	I work hard to do the tasks designated to me.	.70 (.02)		.32 (.04)
Item 9	I execute my tasks foreseeing their results.	.78 (.02)		.32 (.04)
Item 10	I seize opportunities that can improve my results at work.	.72 (.02)	n.s.	
Goodness of fit				
χ^2 (df) = 201.3 (28)				
TLI = 0.98	CFI = 0.99			
RMSEA (90% CI) = 0.06 (0.05 − 0.07)				
	CR	.91	.41	.23
	ωh	.88	.02	.02

Notes. n.s. = not statistically significant ($p<0.05$); The empty cells indicate that the factor loading was constrained to zero; CR = Composite Reliability; ωh = Hierarchical Omega.

The single factor model did not show adequate goodness of fit indicators either. Nevertheless, the borderline fitness indicators point to the possibility of a dominant overall dimension. Furthermore, the adjustment of the single and two-dimensional models (both borderline), as well as the high correlations between the factors are indicative that the factorial model can be more complex. Alternatively, the bifactor model, with one general and two specific dimensions, fitted adequately to the data. The factor loadings of the general dimension ranged from 0.59 to 0.82 ($M = 0.71$), and the factor loadings of the specific dimensions ranged from 0.06 to 0.67 ($M = 0.33$), which can be verified in Table 2.

Regarding reliability, the scores of General Factor, Task and Context showed Composite Reliability (CR) equal to .91, .41, and .23, respectively. The Hierarchical Omega (ω_h) were .88, .02, and .02, respectively. The small factor loadings of the specific dimensions, as well as their low reliability, point out that the overall dimension is more appropriate. However, the separation theoretically coherent of items into their dimension encouraged testing the bifactor structure in Study 2, as will be presented later.

Discussion

This study aimed to reduce the Self-Assessment Scale of Work Performance, so that it can be applied more quickly, in a shorter time without losing the psychometric characteristics of the original scale. Thus, after the confirmatory factor analysis, the bifactor model showed a better fit to the data.

The final model maintains a general factor (with higher loadings) and two specific dimensions (task and context).

The original and the short scales used the same database; however, the results of the short scale presented discrete differences due to the models used in each version. The results presented in the original scale (Queiroga, 2009) indicated that the instrument has good psychometric fitness coefficients, and the model used was the two-factor model, being one related to the task and another to the context. The technique used in the original model does not permit the simultaneous estimation of an overall dimension and two specific dimensions. The results analyzed in the reduced scale also showed good psychometric adjustments, but the model used was the bifactor. This model offers an advantage for the researcher because it works with the simultaneous estimation between the general dimension and the specific dimensions, as it permits the estimation of general scores, in this case, divided into a general factor and two specific dimensions (task and context). Besides, the specific dimensions are mutually independent, which may facilitate studies that relate these variables to others, presenting a better understanding and control of collinearity.

Thus, it could be verified that the short scale, represented by the bifactor model, is composed of a general factor (with higher loadings) and two specific dimensions related to task and context. This bifactor model accommodates the two-dimensional performance model proposed by Borman and Motowidlo (1997) and revisited by Sonnentag and Frese (2002), as it aims to measure performance as individual behaviors related to task and context.

It should be noted that the bifactor model presents a general factor with high factor loadings. Hypothetically, the general factor may contain genuine content variance as well as variance due to the response bias (Danner et al., 2015; Rammstedt & Farmer, 2013). For this reason, in study 2, we tried to evaluate the structure of the short version by using the control for response style and acquiescence through inverted items and anchoring vignettes.

Study 2

Method

Participants

This stage involved 313 Brazilian workers. Among the answers obtained in the application of the short scale, 67% were received online using the tool Qualtrics, while the other answers were collected face-to-face. The mean age of the respondents was 38 years (SD = 12), 53% of the respondents were female, 49% were married, and 57% had no children. Regarding the educational level, 49% of the participants held a postgraduate degree, 31% had a monthly income ranging from R$ 2,900 to 7,249.99 (Brazilian currency). Concerning the current time on the job, the mean was 7.3 years ($SD = 8.3$); and the mean total work experience was 13.3 years ($SD = 10$ years). Two participants were excluded because they presented the same response pattern (same answers in the questionnaires, inverted items, and vignettes).

Instrument

The performance assessment instrument used was the short version of the Self-Assessment Scale of Job Performance (SJoP). This instrument was presented in the method of Study 1.

We control the response bias and acquiescence using four inverted items (two related to the task and two to the context) besides anchoring vignettes. A study by Primi et al. (2016) indicated that the anchoring vignettes help in the control of the self-style response. Thus, for this study, three vignettes were created, describing examples of individuals with high, medium, and low job performance. An example of a vignette is: "The employee Sebastian, at work, does not seem to like what he does and is not interested in things that happen in the company, he is not very motivated and does not show engagement in the accomplishment of his tasks. He is not dedicated to improving the status of his goals. How high do you rate Sebastian's job performance?" The participant should then rank the character in the vignette on a five-point performance scale, ranging from 1 (very low) to 5 (very high).

Procedure

The first procedure, to control for response bias, involves recoding the scores according to the vignette scores. Each vignette was answered using the same Likert scale as the items. The vignettes work as three thresholds for recoding raw items' scores. Thus, the score assigned by an individual to the lowest vignette, for example, indicates the inferior threshold of the new score. Hence, an item with a raw score lower than this threshold should be coding as 1 (value 1 is the lowest score on the new scale, and indicates a lower score than the behaviors described in the first vignette). Considering the original Likert scale has five categories, the raw scores were transformed into seven points: 1 = if raw score was below the first vignette (or first threshold); 2 = raw score equal to the first vignette; 3 = raw score between the first and the second vignettes; 4 = raw score equal to the second vignette; 5 = raw score between the second and third vignettes; 6 = raw score equal to the third vignette; 7 = raw score above the third vignette (further details on the procedure are available in Primi et al. (2016)).

The second procedure, to control for acquiescence, involved the ipsatization of the scores (Soto et al., 2008; Ten Berge, 1999). We first calculated a response style score for each participant, averaging the positive key items and their respective negative items. In this sense, if the participant was not acquiescent, this average should be around 3, which means perfect opposite answers for positive and negative pair of items (for instance, when the participant rates the positive item as 5, he should endorse the category 1 for the negative item; and both answers average 3). However, the average of positive and negative items above three indicates positive acquiescence, and likely a biased score. The average (of positive and negative items) was subtracted from the raw scores (i.e. new score = raw score – average response style) to partial out the acquiescence.

Acquiescence was also tested through a Random Intercept Model (Maydeu-Olivares & Coffman, 2006), using four pairs of positive and negative items (i.e., eight items in total). We set a general factor and an additional method factor (or a Random Intercept) related to response bias, fixing all loadings as equal to 1 (including the negative items). Then the random intercept can capture the idiosyncrasy of the answers. We used only these eight items aiming to keep the balance of positive and negative items; otherwise, the random intercept could super estimate the bias and "steal" part of the true content variance of the general factor.

We performed a Confirmatory Factor Analysis through Structural Equations Modeling to verify the structure of the scale. For this purpose, the same parameters and indicators described in Study 1 were adopted.

Results

A confirmatory factor analysis of the bifactor model was performed without control for response bias, which fitted to the data. The factor loadings of the general factor varied

from .52 to .71 (M = .64), and the factor loadings of the specific dimensions ranged from .16 to .64 (M = .46). The task dimension was composed of two items with specific load-

ings, and the context dimension encompassed four items (Table 3).

Table 3. Bifactor model without controlling for response bias, with control for group bias (vignettes) and acquiescence (inverted items).

Items	Without control for response bias			Recoded items based on vignettes			Recoded items based on inverted items (aquiescence)		
	General Factor	Task	Context	General Factor	Task	Context	General Factor	Task	Context
Item 1	.71 (.03)	n.s.		0.79 (0.02)	n.s.		0.71 (0.03)	n.s.	
Item 2	.69 (.04)		.16 (.06)	0.76 (0.03)		0.23 (0.05)	0.70 (0.03)		0.15 (0.05)
Item 3	.70 (.04)	n.s.		0.75 (0.03)	n.s.		0.74 (0.03)	n.s.	
Item 4	.52 (.05)	.64 (.03)		0.66 (0.03)	0.53 (0.03)		0.57 (0.04)	0.61 (0.03)	
Item 5	.58 (.04)	.64 (.03)		0.72 (0.03)	0.53 (0.03)		0.64 (0.03)	0.61 (0.03)	
Item 6	.68 (.04)		.61 (.08)	0.74 (0.03)		0.53 (0.05)	0.61 (0.04)		0.71 (0.09)
Item 7	.61 (.04)		.43 (.07)	0.72 (0.03)		0.38 (0.05)	0.60 (0.04)		0.38 (0.06)
Item 8	.69 (.04)	n.s.		0.77 (0.03)	n.s.		0.68 (0.03)	n.s.	
Item 9	.58 (.03)		n.s.	0.70 (0.03)		n.s.	0.47 (0.04)		n.s.
Item 10	.67 (.03)		.25 (.05)	0.76 (0.03)		0.26 (0.04)	0.65 (0.04)		0.21 (0.04)
Goodness of fit									
χ² (df) = 85.5 (30)				χ² (df) = 101.17 (30)			χ² (df) = 129.82 (30)		
TLI = .97	CFI = .98			TLI = .98	CFI = .98		TLI = .95	CFI = .97	
RMSEA (90% CI) = .08 (.06-.10)				RMSEA (90% CI) = .09 (.07-.11)			RMSEA (90% CI) = .11 (.09-.12)		
CR	.88	.58	.34	0.92	0.44	0.34	0.87	0.54	0.40
ωh	.85	.04	.03	0.90	0.02	0.02	0.84	0.03	0.03

Notes. The coefficients between parentheses are standard estimation errors; n.s. = not statistically significant (p < .05); The empty cells indicate that the factor loading was constrained to zero; CR = Composite Reliability; ωh = Hierarchical Omega.

In the analysis of the vignettes, two cases were withdrawn, as they presented the same answers for all vignettes and items. Thus, the analysis consisted of 301 respondents. Concerning the anchoring vignettes, 84.7% of the participants (N = 255) showed answers sorted as expected (i.e., they rated lower for the first vignette, median scores for the second vignette and the highest scores for the third vignette). On the other hand, 13% of the respondents indicated a tied response for two different vignettes (i.e., they rated two vignettes with the same value). Moreover, 2% presented inverted answers for two vignettes, that is, responses with a violation.

Not all participants' responses with violations or ties in the vignettes involved recoding problems; for example, a participant who answered one for the lowest vignette, one for the intermediate and two for the highest vignette violates the recoding rules only for score one. For this example, score two would be recoded as six, and scores three, four, and five would be recoded as seven. In this sense, we tried to evaluate the percentage of responses, specific in the present study, that violated the recoding rules. The occurrence was in only 96 responses (87 for tied vignettes and 9 for order violations), which represent 3.18% of the participants' total responses. We also tried to analyze the models without violations or ties in the vignettes, but we observed no improvement in the goodness of fit, nor factor loadings. In summary,

these results indicate that the vignettes constructed are adequate for this study.

The results of the confirmatory factor analysis for the bifactor model using the anchoring vignettes (Table 3) indicated that the model fitted well to the data (CFI = .98 and TLI=.98). The factor loadings have improved, which corroborates hypothesis 1. The factor loadings of the general factor ranged from .66 to .79 with (M = .73), while the factor loadings of the specific dimensions ranged from .23 to .53 and (M = .41). In the task dimension, three items had non-significant loadings, and in the context dimension, one item had a non-significant loading.

A confirmatory factor analysis of the bifactor model with inverted items was executed to control for acquiescence. The ipsatized scores were discretized in ordinal categories to maintain comparability with the previous models. The model presented slightly worse goodness of fit than the previous models. The factor loadings of the specific dimensions were slightly higher than in the vignette models. Still, these loadings were similar to the model for the raw scores (without controlling for response bias). Besides, the loadings of the general factor did not present higher coefficients than those of the other models. Thus, the control for the acquiescence did not improve the factorial solution of the model. Hypothesis 2 of this study was not confirmed.

Although the response bias did not clear the factorial solution, the loadings of the general dimension might be artifi-

cially large due to the response bias instead of real content variance. To test this hypothesis, we set a random intercept model considering only the eight balanced items (i.e., four pairs of positive and negative items), for which the results are in Table 4. In this model, acquiescence, estimated through a random intercept variable, accounts for only 4% of the variance (i.e., $.16^2$). Furthermore, the loadings of the general factor showed slight differences (almost negligible)

between the unidimensional and the random intercept model. The results of the random intercept model also support the rejection of hypothesis 2. Indeed the acquiescence did not consistently bias the factor loadings. As we were using pairs of opposite items, but assessing the same content, we also expected some residual correlations between items. Indeed, two pairs showed significant correlations ($r = .46$ and $.52$), which were freely estimated to adjust the model.

Table 4. Random intercept model with balanced positive and negative worded items and a unidimensional model without controlling the response bias.

Items	Random Intercept Model		Unidimensional
	General Factor	Random Intercept *	
Item 1	.65 (.04)	.16	.66 (.04)
Item 6	.69 (.04)	.16	.69 (.04)
Item 8	.72 (.04)	.16	.72 (.04)
Item 9	.47 (.04)	.16	.48 (.04)
Item 11	-.56 (.04)	.16	-.57 (.04)
Item 12	-.58 (.04)	.16	-.57 (.04)
Item 13	-.68 (.04)	.16	-.65 (.04)
Item 14	-.70 (.05)	.16	-.68 (.05)
Goodness of fit			
χ^2 (df) = 44.98 (17)			χ^2 (df) = 53.69 (18)
TLI = .97 CFI = .98			TLI = .97 CFI = .98
RMSEA (90% CI) = .07 (.05-.10)			RMSEA (90% CI) = .08 (.06-.11)
CR	.84		0.84

Notes. The coefficients between parentheses are standard estimation errors;
* fixed parameter as equal for all items including those negative worded.

Regarding reliability, scores of the model without control for response bias showed Composite Reliability (CR) equal to .88, .58, and .34, respectively, for the General Factor, Task, and Context. The Hierarchical Omega (ωh) were .85, .04, and .03. Controlling for the group bias through vignettes, the CR were .92, .44, and .34; and the ωh were .90, .02, and .02. Controlling for the acquiescence bias through inverted items, the CR were .87, .54, and .40; and the ωh were .84, .03, and .03.

It was possible to verify the occurrence of a slight increase in the reliability in the general factor when using the vignettes control; this fact can be explained by the increase of the variance caused by the process of recoding, which

changes from a scale of five points to seven points. Regarding the control of acquiescence, the accuracy did not show improvement.

In general, the accuracy is high for the general dimension and low for specific dimensions. This fact is common in bifactor models since most of the variance is usually attributed to the general dimension, which after controlling for the variance of this general dimension, there is little remaining covariance to be explained by the specific dimensions. Consequently, in bifactor models, the reliability of the scores of the specific dimensions tends to be lower. The Table 5 compare the full and the short version.

Table 5. Comparison between the items of the original version and the reduced version and the dimensions evaluated.

Items	Description	Full Version	Short Version
Item 1	I perform hard tasks properly.	♦	♦
Item 2	I try to update my technical knowledge to do my job.	•	•
Item 3	I do my job according to what the organization expects from me.	•	•
Item 4	I plan the execution of my job by defining actions, deadlines and priorities.	•	•
Item 5	I plan actions according to my tasks and organizational routines.	•	•
Item 6	I take initiatives to improve my results at work.	•	•
Item 7	I seek new solutions for problems that may come up in my job.	•	•
Item 8	I work hard to do the tasks designated to me.	•	•
Item 9	I execute my tasks foreseeing their results.	•	•
Item 10	I seize opportunities that can improve my results at work.	♦	♦
Item 11	I execute my tasks according its deadlines.	•	-
Item 12	I execute my job according organizational structure and polices.	•	-
Item 13	The performance of my work contributes to the achievement of the Organization's mission and objectives.	•	-
Item 14	I do my job with economy of resources (material, financial and human).	•	-
Item 15	I contribute with alternatives for solving problems and improving the Organization's processes.	•	-
Item 16	I establish contact with other people or teams to achieve organizational goals.	•	-

Items	Description	Full Version	Short Version
Item 17	I adapt my routine to changes in the Organization's goals.	•	-
Item 18	I resolve questions from my colleagues when asked.	•	-
Item 19	I properly perform routine tasks.	♦	-
Item 20	I use my theoretical / practical knowledge to carry out my work.	♦	-

Note: ♦ Task performance; • Context performance

Comparing the original 20 items scales with the short version, we can see both dimensions initially provided are adequately covered in the short version. The contents of the deleted items are covered by items we kept in the short version.

Discussion

This study aimed to confirm the internal structure of the reduced self-assessment scale of job performance (Queiroga, 2009), and to evaluate the effect of response bias (group bias) and acquiescence of the scores on the scale structure. The results confirm the structure of the model of the reduced scale version, presenting one general factor and two specific dimensions (task and context). It is emphasized that the main strength of this model is the general dimension, as it has higher factor loadings when compared to the specific dimensions. Moreover, the bifactor model also presents satisfactory goodness of fit indices. Therefore, these results are in line with the theory by Sonnentag and Frese (2002) regarding the task and context performance, which supported the construction of the original instrument. We also must highlight the estimations of the subjects' scores for a specific factor should be avoided due to the unreliability.

As practical recommendations, it is important to point out that the scale can be better used if all ten items are applied, and a general score is considered. Partialling out the variance of the specific factors might clear out the scores of the general factor, and then, the procedure can be suggested. However, an overall average of the full scale may be recommended, especially if the purpose is to research the perception of job performance. It is important to note that the theoretical structure of the scale has not been altered (see Table 5), that is, no factor has emerged as a result of a new grouping of items. In fact, it was observed that, even for the reduced version, the two dimensions originally proposed are present. However, the short version was more robust in the one-dimensional structure, which is why the recommendation to use this option is recommended.

We used the anchoring vignettes and inverted items to control for the response bias, which is acknowledged as a problem in self-reported scales (Primi et al., 2016; Soto et al., 2008). The results showed that the use of the vignettes had a positive impact on the factor loadings of the general dimension, which confirms hypothesis 1. These results are in line with the studies by Primi et al. (2016), who also verified in their experiments that the use of anchoring vignettes, to control for response bias, improved the factor loadings and reliability of the scores.

It should be noted that the increase in factor loadings is more pronounced in the general dimension than in the specific ones. In this sense, the general dimension is more than a mere consequence of the response bias (according to the hypothesis we made on Study 1 conclusions). Thus, after controlling for effects of personal styles, the content of the items seems to be based on general conditions for carrying out the job activities in a committed way. Besides, the items for the "task" dimension had practically no significant loadings in the specific dimension. The only characteristic of the items related to the task factor that has a specific variance, in addition to the general factor, is related to the planning. In this sense, the "task" dimension, even after controlling for response bias, seems embedded in the general dimension. In other words, the theoretical content of the task, in this case, seems to correspond to the general understanding of job performance. This result resembles the research findings which report the importance of performing tasks corresponding to the function in question to achieve effective performance (Coelho Junior & Borges-Andrade, 2011; Obeidat Shatha et al., 2016; Warr & Nielsen, 2018). In this context, the employees will be able to execute ways to maximize their abilities to perform their functions. Thus, it can be verified that, if the individual presents proficiency in the tasks he/she performs, he/she certainly performs well what he/she is meant to do. On the other hand, even after controlling for the response style, the context dimension continues to present moderate factor loadings. Thus, the instrument items presented specific context contents, which are not fully explained by the general performance. Therefore, individuals can present good strategic behavior, have good social, psychological, and interpersonal relationships, that is, present a contextual performance independently of the overall performance desired by the organization.

Another way to control for the response bias in this study was the use the inverted items related to the task and the context. However, we verified that the use of this procedure did not improve the model beyond the anchoring vignettes. The random intercept model was set up with only eight balanced items and yielded the same conclusion: loadings of the general factor were not consistently biased by acquiescence. Results also point out the strength of the general factor. Thus, hypothesis 2 that the control for acquiescence through inverted items would have a positive impact on the scale structure could not be confirmed. In this sense, the acquiescent style may be very homogeneous in the sample studied or, simply, may not exert significant influence on the job performance scores. Another hypothesis refers to the quality of the inverted items, and the formulation may have

also changed the descriptive content of the items and not just how they were keyed (from positive to negative). In this sense, a future study may test other formulations of inverted items.

Regarding the limitations of the study, the use of a convenience sample ends up limiting the generalization of the results obtained. Another limitation was related to the size of the face-to-face sample in Study 2 (N = 104), which hampers an invariance analysis of the item parameters between the online and the face-to-face application. Besides, this study did not verify the relation between professional performance and performance results. Furthermore, we did not investigate the influence of other sources of response bias, like social desirability. Considering the lack of studies on the impact of social desirability, we do not recommend using the short scale on assessing high stakes groups.

Conclusion

This study was aimed at reducing the Self-Assessment Scale of Job Performance. The results indicated evidence of a unifactorial model.

Future studies could whether the performance scores produced on this scale are truly related as work results (car sales, for example). Maybe it would also increase the understanding of the nomological network of the construct. Besides, future research could be carried out to increase the size of the face-to-face and online samples, so that an invariance analysis can be performed. Moreover, future research could aim at other sources of response bias, like social desirability.

Overall, we make available a short scale, with only ten items (related to task and context), which can be applied more quickly. Besides, the internal structure evidence emphasizes the adequacy of using the scale in future research. Although there are other bias control methods (example: a balanced set of positively and negatively keyed items on the scale), this study chose to study the anchoring vignettes and inverted items to allow comparison with results using similar procedures in the Brazilian context (Primi et al., 2016). Thus, this study advances previous research by trying to control for response biases, which are common problems in the answers to self-reported scales, in that they do not reliably portray the respondent's reality.

References

Bing, M. N., Davison, H. K., Minor, I., Novicevic, M. M., & Frink, D. D. (2011). The prediction of task and contextual performance by political skill: A meta-analysis and moderator test [Article]. Journal of Vocational Behavior, 79(2), 563-577. https://doi.org/10.1016/j.jvb.2011.02.006

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. Human Performance, 10(2), 99-109. https://doi.org/10.1207/s15327043hup1002_3

Bowling, N. A., Khazon, S., Meyer, R. D., & Burrus, C. J. (2015). Situational Strength as a Moderator of the Relationship Between Job Satisfaction and Job Performance: A Meta-Analytic Examina-tion. Journal of Business and Psychology, 30, 89-104. https://doi.org/10.1007/s10869-013-9340-7

Brandão, H. P., Borges-Andrade, J. E., & Guimarães, T. d. A. (2012). Desempenho organizacional e suas relações com competências gerenciais, suporte organizacional e treinamento [Organizational performance and its relations with management competencies, organizational support and training]. Revista de Administração (São Paulo), 47, 523-539. https://doi.org/10.5700/rausp1056

Byrne, B. M. (2013). Structural equation modeling with mplus: Basic concepts, applications, and programming. Taylor & Francis. https://books.google.com.br/books?id=8vHqQH5VxBIC

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 and organizational psychology (2nd ed., Vol. 1, pp. 687-732). Consulting Psychologists Press.

Campbell, J. P. (2012). Behavior, performance, and effectiveness in the twenty-first century. In S. W. J. Kozlowski (Ed.), The Oxford handbook of organizational psychology (Vol. 1, pp. 159-194). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199928309.013.0006

Campbell, J. P., & Wiernik, B. M. (2015). The Modeling and Assessment of Work Performance. Annual Review of Organizational Psychology and Organizational Behavior, 2, 47-74. https://doi.org/10.1146/annurev-orgpsych-032414-111427

Coelho Junior, F. A., & Borges-Andrade, J. E. (2011). Efeitos de variáveis individuais e contextuais sobre desempenho individual no trabalho [Effects of individual and contextual variables on individual job performance]. Estudos de Psicologia (Natal), 16, 111-120. https://doi.org/10.1590/S1413-294X2011000200001

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. Journal of Research in Personality, 57, 119-130. https://doi.org/10.1016/j.jrp.2015.05.004

Edwards, B. D., Bell, S. T., Arthur, W., & Decuir, A. D. (2008). Relationships between facets of job satisfaction and task and contextual performance. Applied Psychology-an International Review-Psychologie Appliquee-Revue Internationale, 57, 441-465. https://doi.org/10.1111/j.1464-0597.2008.00328.x

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3(4), 424-453. https://doi.org/10.1037//1082-989x.3.4.424

Huang, W.-R., & Su, C.-H. (2016). The mediating role of job satisfaction in the relationship between job training satisfaction and turnover intentions. Industrial and Commercial Training, 48(1), 42-52. https://doi.org/10.1108/ICT-04-2015-0029

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions. (pp. 3-90). Jossey-Bass.

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. Psychological Methods, 11(4), 344-362. https://doi.org/10.1037/1082-989X.11.4.344

Obeidat Shatha, M., Mitchell, R., & Bray, M. (2016). The link between high performance work practices and organizational performance: Empirically validating the conceptualization of HPWP according to the AMO model. Employee Relations, 38(4), 578-595. https://doi.org/10.1108/ER-08-2015-0163

Paula, A. P. V. d., & Queiroga, F. (2015). Satisfação no trabalho e clima organizacional: a relação com autoavaliações de desempenho [Job satisfaction and organizational climate; the relation with performance self-assessment]. Revista Psicologia Organizações e Trabalho, 15, 362-373. https://doi.org/10.17652/rpot/2015.4.478

Primi, R., da Silva, I. C. R., Rodrigues, P., Muniz, M., & Almeida, L. S. (2013). The use of the bi-factor model to test the uni-dimensionality of a battery of reasoning tests. Psicothema, 25(1), 115-122. https://psycnet.apa.org/record/2013-02569-019

Primi, R., Zanon, C., Santos, D., Fruyt, F. D., & John, O. P. (2016). Anchoring Vignettes: Can They Make Adolescent Self-Reports of Social-Emotional Skills More Reliable, Discriminant, and Criterion-Valid? European Journal of Psychological Assessment, 32(1), 39-51. https://doi.org/10.1027/1015-5759/a000336

Queiroga, F. (2009). Seleção de pessoas e desempenho no trabalho: Um estudo sobre a validade preditiva dos testes de conhecimentos [Staff selection and job performance: A study on the predictive validity of knowledge tests] [Doutorado Tese, Unb]. Brasília. http://repositorio.unb.br/bitstream/10482/8437/1/2009_FabianaQue iroga.pdf

Queiroga, F., Borges-Andrade, J. E., & Coelho Junior, F. A. (2015). Desempenho no trabalho: Escala de avaliação geral por meio de autopercepções [Job performance: General assessment scale through self-perceptions]. In K. Puente-Palacios & A. d. L. A. Peixoto (Eds.), Ferramentas de diagnóstico para organizações e trabalho: Um olhar a partir da psicologia [Diagnostic tools for organizations and work: A psychological perspective]. Artmed Editora.

Rammstedt, B., & Farmer, R. F. (2013). The Impact of Acquiescence on the Evaluation of Personality Structure. Psychological assessment, 25(4), 1137-1145. https://doi.org/10.1037/a0033323

Sonnentag, S., & Frese, M. (2002). Performance concepts and performance theory. In S. Sonnentag (Ed.), Psychological management of Individual performance. Wiley.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. Journal of Personality and Social Psychology, 94(4), 718-737. https://doi.org/10.1037/0022-3514.94.4.718

Ten Berge, J. M. F. (1999). A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization. Multivariate Behavioral Research, 34(1), 89-102. https://doi.org/10.1207/s15327906mbr3401_4

Valentini, F., & Damásio, B. F. (2016). Variância Média Extraída e Confiabilidade Composta: Indicadores de Precisão [Average Variance Extracted and Composite Reliability: Reliabili-ty Coefficients]. Psicologia: Teoria e Pesquisa [Psychology: Theory and Research], 32, 1-7. https://doi.org/10.1590/0102-3772e322225

Warr, P., & Nielsen, K. (2018). Wellbeing and work performance. In E. Diener, S. Oishi, & L. Tay (Eds.), Handbook of well-being. DEF Publishers. https://doi.org/nobascholar.com