

Team Performance: Evidence for Validity of a Measure

Katia Puente-Palacios – Universidade de Brasília, Brasília, Brasil

Maria do Carmo Fernandes Martins – Universidade Metodista de São Paulo, São Paulo, Brasil

Salvatore Palumbo – Tribunal de Contas da União, São Paulo, Brasil

Abstract

This study aimed to obtain validity evidences of a teamwork performance scale. Team performance was understood as a meso-level characteristic, resulting of an emerging process. Due to that understanding the proposed instrument should take that aspect into consideration. The empirical data were collected from 276 Ecuadorian teachers organized in 70 educational teams. They answered nine sentences of the scale. Results of exploratory factor analysis showed a unifactorial solution explaining 65.84% of the variance. The measure also has adequate values of reliability (Cronbach Alpha = .93). In addition to these analyses, patterns of variance within and between the groups were verified. The results showed that the variance at the individual level was small when answers of team members were analyzed and was significant when teams were compared. We consider it is important that additional studies be performed in order to identify stability of the factor solution.

Key words: job performance; assessment scale; teamwork

Desempenho de equipes: evidências de validade de uma medida

Resumo

O objetivo deste estudo foi desenvolver e verificar evidências de validade de uma medida que permita aferir o desempenho de equipes de trabalho. O desempenho foi entendido como atributo do nível meso, portanto, relativo às equipes, que surge mediante um processo de emersão e, desse modo, a medida deve considerar essa especificidade. Para a realização empírica da pesquisa contou-se com respostas de 276 professores equatorianos organizados em 70 equipes pedagógicas, que responderam a nove afirmativas presentes no instrumento. Os resultados de análise fatorial exploratória mostraram a solução unifatorial como a mais satisfatória, a qual captura 65,84% da variância do fenômeno e possui valores adequados de confiabilidade (*alpha* de Cronbach de 0,93). Análises de padrões de variância dentro e entre grupos revelaram que a medida permite capturar um fenômeno do nível meso. Recomenda-se a realização de outras pesquisas que verifiquem a estabilidade desta solução.

Palavras-chave: desempenho no trabalho; escala de avaliação; equipes de trabalho.

Rendimiento de equipos: evidencias de validez de una medida

Resumen

El objetivo de este estudio fue desarrollar y verificar evidencias de validez de una medida que permita la medición del rendimiento de equipos de trabajo. El rendimiento fue entendido como un atributo del nivel meso, o sea, relativo a los equipos, que surge mediante un proceso de emersión, y de ese modo la medida propuesta debería llevar en consideración ese aspecto. Para la realización empírica de la investigación se contó con las respuestas dadas por 276 profesores equatorianos organizados en 70 equipos pedagógicos, que respondieron a nueve afirmativas contenidas en el instrumento. Los resultados del análisis factorial exploratorio mostraron la solución unifatorial como la más satisfactoria que capturó 65,84% de la varianza del fenómeno y posee valores adecuados de confiabilidad (*alpha* de Cronbach de 0,93). Análisis de patrones de varianza dentro y entre grupos revelaron que la medida permite capturar un fenómeno del nivel meso. Por fin, se recomienda la realización de otras investigaciones que verifiquen la estabilidad de esta solución.

Palabras-clave: rendimiento en el trabajo; escala de evaluación; equipos de trabajo.

Work teams are performance units organized to promote organizational effectiveness; thus, their presence reveals interest in agile response to the current demands of the working world. The emphasis on understanding what these teams are and what are their main attributes has been appearing in various publications in the area. In this regard, Cohen and Bailey (1997) point out that, since the proposed comprehensive structure of the elements involved in the operation of teams, suggested by MacGrath and dated 1964, a variety of efforts have been undertaken, seeking to

understand both their operation and the results that can be expected.

The proposed structure mentioned above considers the presence of three key elements: inputs, processes, and outputs. The original model was modified on the basis of criticisms received, and as a result, the process element was divided into two parts, with the first keeping its original name (processes) and the second receiving the designation of emergent states (Mathieu, Maynard, Rapp, & Gilson, 2008). Despite the time elapsed since its appearance, as well as the

modifications applied, more recent publications demonstrate the utility of this model (Bell, 2007; Callea et al. 2014; Puente-Palacios & González-Romá, 2013).

The relevance of the above proposal lies also in the fact of presenting a logical way of organizing the aspects involved in the operation of teams, as well as in the prominence given to the different types of possible results, which constitute the primary consequence sought with their implementation. Briefly, the inputs can be described as initial attributes, or those essential to the function of the teams. Among them are found characteristics of the organization, of the team as a whole, and of its members. The processes are viewed as the transformations that occur over the lifespan of the team, and may be related both to the task and to the relationships maintained between the members. Finally, the outputs are the consequences of the experience of working together, including both those desired and expected, as well as those unwanted and even avoided, and may be related to the members, to the team as a whole, or to the organization (Mathieu et al., 2008). The object of interest for this manuscript is desired output or expected results.

Discussing desired results, Nadler, Hackman and Lawler (1979), and later Hackman (1987), adopted the term effectiveness, and explained that this involves three criteria. The first refers to the delivery of the product, which must meet the time and quality conditions set by customers, whether internal or external to the organization. According to Hackman, this is not limited to concrete or hard indicators, because the really important aspect is the client assessment of the product. Thus, judgments made about the team's work form the core of this criterion.

The second criterion is the favorable affective result arising from the shared work experience, which can be operationalized in the form of satisfaction with the team and with the work performed. The third criterion, in turn, relates to the so-called viability or survival of the team, which can be described as the ability of the team members to embrace a new experience of working together after the experience of performing a task that required a combination of individual and collective efforts.

With regard to the names adopted to refer to performance results, Brodbeck (1996), after a review of the empirical research, warns against the use of various terms that bring added confusion to this field of knowledge. He mentions that the same criterion can be named productivity, effectiveness, or even performance.

Similar criticism was made by Puente-Palacios and González-Romá (2013) who, following the thinking of the first author, point out specificities of each term. Thus, it is mandatory specify a proper definition of the phenomenon, as a necessary requirement for advancement of knowledge.

In theorizing about the model that organizes the elements involved in the operation of the teams, Mathieu, Maynard, Rapp and Gilson, (2008) mention that much more is known about the first two (inputs and processes) and less about the third (outputs). They attribute this, in large part, to the measurement instruments, which are described as scarce, and to the fact that organizations tend to develop customized scales that capture only their own reality. Focusing on the specificity that characterizes these measures, McMillan, Entin, Morley, and Bennett Jr. (2013) discuss the existence of instruments tailored to specific scenarios, and following the central trend of this field, propose a measure for use in military environments, applicable to teams composed of four members, F16 aircraft pilots. From these findings, it is concluded that the development of a measure that facilitates team performance assessment, for cross-organizational use, applicable to teams that perform many different tasks, is an important contribution.

In relation to what must be understood as team performance, Brodbeck (1996) establish it is a set of behaviors required for work goals to be achieved. Salas, Cooke, and Rosen (2008), in turn, conceptualize it more as a process than as a result, and argue that it encompasses cognition, attitudes, and behaviors, which act in an interrelated and interdependent manner. Thus, authors of the area defend that team results are compound of behaviors, motivations, and attitudes shared by the members which arise during the lifespan of the team.

The discussion concerning the distinction between processes and results of team work is not without controversy, as is the case with many phenomena in the organizational psychology field. Mathieu et al. (2008), based on the interpretations of Beal, Cohen, Burke, and McLendon (2003), argue the difference between performance behaviors and performance results. Performance behaviors are described as actions relevant to reaching the goal, while performance results would be their consequences. The study by Beal et al. (2003) meta-analytical in nature, sets out from the principle that cohesion is the result of the occurrence of certain group processes, thus

being an output variable. Therefore, it should show significant correlations with other team result criteria. However, data analysis showed that cohesion had stronger correlations with performance when it was operationalized as process behaviors, compared to correlations observed when operationalized as behaviors of result. From these findings, the authors draw conclusions about the differences resulting from the way performance is understood and operationalized, and point out the need for proper definition of the nature of what will effectively be evaluated: a group process or a result of the group process.

Brannick and Prince (2009), discussing this issue, also indicate that among the collective or team processes, there is a wide range of those that are included in team performance evaluations. They mention coordination and communication as those usually focused upon in such measurements. They also point to the difficulty of defining which should be the object of diagnosis and emphasize that, in many cases, this definition stems from the purpose of the assessment. That is, if used, for example, for decisions on training actions, the processes focused upon should probably be those aligned with individual and collective competencies. But if the purpose is to offer feedback to the team, it might be appropriate to assess attributes such as coordination, communication, and interaction processes, just to name a few.

The distinction between performance, seen as a process (behaviors) or as a result, also appears in the literature that discusses individual performance. Sonnentag (2002), a prolific author in this field of knowledge, affirms that in the first case, these are actions taken by the worker that help ensure the task can be performed. On the other side, the results are described as consequences or indicators of the work performed by the individual. Thus, it is observed that although the focus of this article is team performance, not individual performance, the construct proves to have a similar nature at both levels. The differential aspect of its manifestation at the collective level (in this study it is called team performance) is in the sharing observed between members, which results from individual contributions that are transformed and joined, giving rise to a characteristic of the team.

With this scenario in mind, the phenomenon effectively focused upon in this article constitutes one of the consequences of group processes, i.e., it is an output criterion. Thus, the proposed measure addresses the team's work results and not the processes involved

in completing the task. Following the theories by Hackman (1987), it is understood that the assessments made of the work carried out constitute the aspect of central importance. For this reason, collective performance descriptors are defined, and will be evaluated on the basis of judgments.

It is also important to note that related to team performance, what is the main aspect emphasized is the fact that it is shared. This is because the performance of the meso level, or of a collectivity, results from a process of emergence in which various contributions made by the members are integrated and combined, in a dynamic and complex way. Defining this level as a target of interest demands, in addition to adopting theoretical perspectives that take it as a collective attribute, compliance with methodological requirements. In this regard González-Romá, Fortes-Ferreira, and Peiró (2009) highlight the need for collective performance measurement to demonstrate the sharing and emergence in which such constructs (at the meso level) are sustained. While Coultas, Driskell, Burke, and Salas (2014), specifically focusing on the processes of emergence, indicate the need to observe the theoretical nature of the construct, the requirements linked to the development of the measurement, and the required analytical strategies.

Describing the specificity that should characterize the tools for measuring group phenomena, Puente-Palacios and Borba (2009), in turn, affirm that data collection at the individual level and later aggregation of scores to compute a representative value or score for the teams is not sufficient. It is essential, according to the authors, that the questions of the scale address properties of the team, which in this case would be its performance.

The requirement to define different strategies for measuring collective characteristics or behaviors is supported by the fact that many of them arise through processes of emergence (Coultas, Driskell, Burke, & Salas, 2014; González-Romá et al., 2009). For Klein and Kozlowski (2000), these are transformations of individual attributes from which shared or collective characteristics originate, which are termed bottom-up relations. This name conveys the idea of emergence of a new phenomenon resulting from changes or shifts in properties existing before only at the individual level. Processes of emergence are also described by Kozlowski, Chao, Grand, Braun, and Kuljanin (2013) as dynamic and interactive phenomena, multilevel in nature.

Regarding the recognition of the nature of the phenomena at the meso level, which arise from the combination of individual attributes, Kozłowski et al. (2013) consider very important because both, the theories and the analytical strategies, complement the organizational behavior field, which is lacking in research that integrates multiple levels.

The discussion concerning the occurrence of phenomena at the meso level and the attention that must be paid during measurement process have been developed since the end of the twentieth century. Chan (1998) defends the existence of theoretical models that adequately represent the diversity of this field and proposes five alternatives referred to as emergence models. Based on them, the author explains the theoretical nature of those events and describes the transformations underlying the dynamic processes of combining individual attributes that give rise to collective properties. The relevance of this proposition has been demonstrated in research with a focus on work teams, conducted inside and outside Brazil, that applies the emergence models proposed by the author (Deshon, Kozłowski, Schmidt, Milner, & Wiechmann, 2004; González-Romá & Hernández, 2014; Puente-Palacios & Borba, 2009; Puente-Palacios, Silva, & Borba, 2015; Priesemuth, Schmienke, Ambrose, & Folger, 2014), which reinforces the decision to adopt it in the research objective of this report.

Of the emergence models proposed by Chan (1998), the one that best addresses the specificities of team performance, seen as a result, is the so-called referent-shift consensus. The author states that this type of emergence applies to phenomena that manifest themselves in a similar manner at the individual and the group level, as in the case of performance. For example, beliefs can be theorized as characteristics of the individual subject, but can also be considered as team characteristics. However, when the concern is to focus on collective beliefs, the measures or instruments should not ask the respondent about what he/she believes. They should ask about the beliefs present in their team. That is, the referent or the stimulus that elicits the subject's response is changed, and must aim for what is shared by the members as a group.

In the case of team performance, understanding that this is an attribute resulting from a process of emergence, the referent-shift model proves to be compatible, because although the individual is asked to evaluate the performance, the instrument's questions should focus on team results and not on the work done

by the individual subject. However, it can be stated that it is the team's assessment of its performance only after verifying that the individual judgments are similar, can be combined, and based on them compose a single indicator.

As for the team's attributes, it is still important to note that not all arise by emergence (Klein & Kozłowski, 2000). There are collective characteristics that can be legitimately captured at that level, as in the case of team size or budget available to it to carry out some project. While Denisi (2000), in referring to the team's performance, adopts the term global properties, and states that certain indicators may be observable, and not result from the combination of individual contributions. It should also be noted that in cases of performance evaluations done by the manager, for example, making general judgments on the results of the work is legitimate. In this case, information will be provided by a single actor who would be authorized to assess performance, but by adopting this strategy the researcher will not be able to detect the appearance or emergence of the collective attribute.

Puente-Palacios and Borba (2009) offer additional information about the requirements concerning the development of measures focused on the meso level. They point out that, as the studies are concerned with group attributes, the individual responses cannot be intuitively taken as evidence of collective behavior. They constitute the starting point that enables identification of the process of emergence, which still must be investigated using analytical strategies. The results of these analyses will demonstrate the level of similarity present in the individual responses and confirm the existence of a collective characteristic or variable.

Even so, it must be recognized that the use of this strategy is a subject of debate, as some authors argue the relevance of adopting alternative mechanisms such as holding consensus meetings as a way to directly obtain group information or collective data (Quigley, Tekleab, & Tesluk, 2007). However, questions can also be encountered regarding the use of this method on the grounds that, during the meetings, group phenomena such as pressure for agreement, the existence of coalitions, and power differences may mean the information provided does not reflect the collective thinking, as well as not allow discovery of the process of emergence (Puente-Palacios & Borba, 2009).

Regarding the use of individual information aggregated to form group scores, Smith-Crowe, Burke, Kouchaki and Signal (2013) report that, in journals from

the Psychology field, such as the *Journal of Applied Psychology* and *Personnel Psychology*, published in 2010, roughly half of the articles used some kind of agreement index among evaluators to compose collective indicators resulting from information collected at the individual level. This evidence demonstrates the relevance of advocating the use of information provided by members, when the phenomenon of interest is a collective property. If similarity of the group members' responses is found, then the phenomenon captured is in fact from the group level.

As briefly described, when it comes to measuring meso-level phenomena, collecting information provided by the members is a first step that must be followed by verification of the similarity of the responses from the members in each group. Once verified, the researcher will have evidence that the measured aspect has a collective nature. However, these necessary steps are not sufficient. This is because differences between groups must still be verified. That is, the similarity of group member responses must be accompanied by differences between groups. If both conditions are present, then the researcher is in fact dealing with a group attribute.

In order to conduct this study, performance was defined as a group result whose measurement can be done based on members' reports. The following section describes the method used in the process of developing the measure, which can be adopted both for assessments via self-reporting (by the team members) and for assessments made by the leader or others (supervisor or client).

Method

Participants

To investigate evidence of the validity of the proposed measure, we used data from a sample of teachers at Ecuadorian educational institutions, organized in teaching teams, whose work goals were eminently collective and who carried out activities characterized by interaction and interdependence. Analyses were conducted on the responses from 276 teachers belonging to 70 educational teams. In general terms, these respondents were primarily male (61.6%), with an education level corresponding to a college degree (38.4%), and some of them (10.5%) had studied Pedagogy. The average age was 42.1 years ($SD = 10.5$), and half of the respondents had at least five years of seniority in their

institution. As for the institutions, a large majority was private (81.2%).

Instrument

To develop the instrument, articles from the area where attributes of individual and team performance were discussed were taken as the starting point (Beal et al., 2003; Brannick & Prince, 2009; Brodbeck, 1996; DeNisi, 2000; Mathieu, et al., 2008; Puente-Palacios & González-Romá, 2013; Salas, et al., 2008; Priesemuth, et al., 2014; Sonnentag, 2002). In this way, twelve initial items describing team result behaviors were drafted. The items were then submitted for evaluation by twelve judges, expert researchers in the field who analyzed the relevance of the items for the construct. These evaluators could also suggest, if they thought necessary, topics not covered by the set of items submitted for inclusion in the measure. As a result of this evaluation process, nine items were considered suitable, as agreement between the judges reached 85%. Three items did not reach this level and, therefore, were eliminated. There were no suggestions for including other items. Thus, with the nine items that were positively assessed, the work teams performance evaluation measure was constructed. The extent of the scale was considered appropriate, according to the judges who participated in this stage, for condensing the central descriptors of team performance, without invading parallel fields, such as that of group processes.

The scale effectively applied to the study participants was composed of nine descriptions of possible results achieved by the teams, which had to be answered on a type Likert agreement scale. The respondents had to assess to what extent the statement corresponded to what their team does, or to the results that it presents. The range of the scale was set at 5 points, where 1 corresponded to 'Totally Disagree' and 5 corresponded to 'Totally Agree'. In observance of the precepts for building meso-level measures (Chan, 1998; Puente-Palacios & Borba, 2009) all items were drafted with a focus on what the team does and not on individual performance.

Procedure

Data collection occurred in a face-to-face setting, using the questionnaire in printed form (paper and pencil) for this purpose. Research team members went to the workplace of the teams, after authorization by the organization, and collected the data. To this end, they first provided information about the content of the research, its voluntary nature, and the fact that

participation did not carry any personal or professional consequences. It is also worth mentioning that the ethical principles governing research with human beings were adopted for conducting the research. Thus, participation was voluntary, it took place only after explaining the research content (presented verbally and in writing), the data collected were handled such that the anonymity of participants was protected, and no harmful consequences resulted from the fact that people chose whether or not to answer the scale's questions. Questionnaires were given to those who agreed to answer them, and after being completed, they were collected.

Data Analysis

Considering that the aim of the study was to develop a measure, the analytical procedures adopted were related to the analysis of the factorability of the data matrix, identifying the appropriate number of factors to retain, as well as investigating the adequacy of the solution obtained, by verifying the internal consistency indices of the factors. Next, the level of similarity of the responses given by the members of each team was also verified, using the average deviation index (AD_{Md}) calculation. Finally, evidence of variance between teams was investigated by calculating the magnitude of the intraclass correlation coefficient (ICC) as well as difference in the means, using ANOVA (one-way). The results from applying this set of strategies are described below.

Results

The analytical strategies adopted were aligned with the purpose of the study, namely, to develop a team performance evaluation measure with satisfactory evidence of validity. Thus, we initially sought to understand the overall pattern of responses, and this investigation revealed that the amount of missing data, by item, did not surpass 3.5%, which is why they were not replaced or treated in any way. We also analyzed the distribution of the responses per item and was verified that, although they did not fit the normal curve, the skewness values did not exceed those indicated by Miles and Shevlin (2001) as problematic (above 2 in absolute values).

With these data demonstrating the relevance of the data matrix for the desired analyses, we proceeded to verification of factorability. The criteria used were: calculating the determinant of the matrix, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy,

significance of Bartlett's test of sphericity, and inspection of the correlation matrix.

The results were satisfactory in that the value of the determinant was small, but different from zero (0.002), Bartlett's test was significant ($p < 0.001$), the KMO reached an adequate value (0.94), and all the items showed significant correlations between them (ranging between 0.48 and 0.73).

Based on these findings, it was considered appropriate to invest efforts in reducing the data matrix to factors. To this end, the Principal Axis Factoring (PAF) method was adopted. The appropriate number of factors to retain was identified by adopting criteria such as eigenvalue equal to or greater than 1, or the Kaiser-Guttman rule, which indicates the maximum number of factors supported by the matrix. The Scree Plot criterion (known as Cattell's scree test) shows, in a visual way, the number of factors present in the matrix. Finally, the theoretical criterion that describes the nature of the phenomenon on which the scale is focused was considered. The results showed that the application of statistical criteria would suggest retaining a single factor (Kaiser-Guttman and Cattell), as shown in Figure 1. As to the theoretical nature of performance, it is emphasized that the aspect in focus is a result and not a processes. Thus, a one-factor solution is considered theoretically satisfactory.

The extraction of the single factor allowed capture of 65.84% of the variance of the measured phenomena, and the nine constituent items participated, showing factor loadings whose values were between 0.70 and 0.87. From the items of the scale, the one that best represents the underlying construct describes the team's performance, highlighting that it successfully meets its work goals, followed by the one that focuses on the quality of products / services produced by the team. Table 1 shows these results.

Once the appropriate number of factors to retain was identified, we proceeded with verifying the reliability of the retained factor, using the value of Cronbach's alpha, and the arithmetic mean of the item-total correlation. The results indicated the adequacy of the retained factor, as the Alpha attained a value of 0.93, while the magnitude of the item-total mean correlation coefficient was 0.76.

After finishing the exploratory factor analysis phase, it was verified the occurrence of emergence process mentioned in the introduction section of this article, since the data were collected from team members, but the focus of investigation was on the teams.

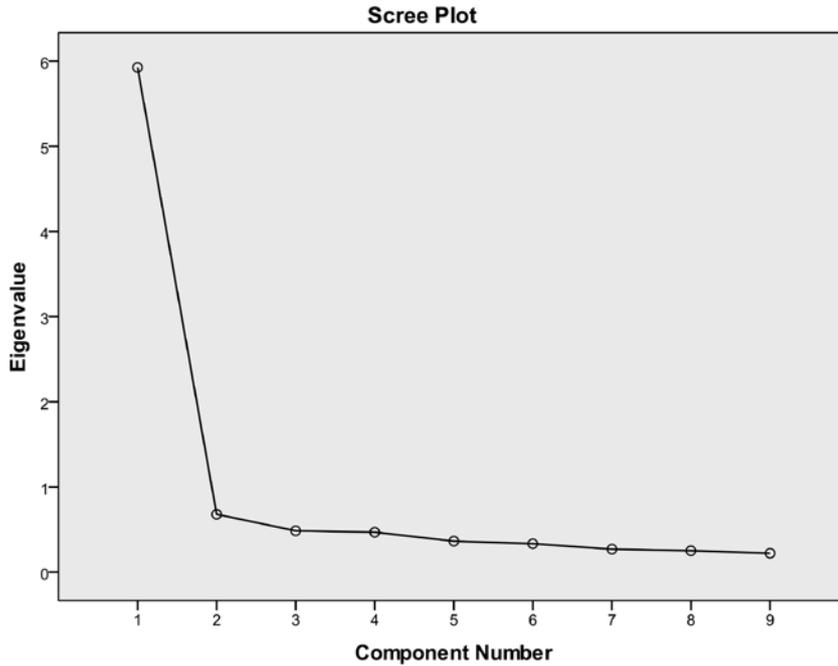


Figure 1. Scree plot, for the performance scale.

Table 1
Factor Loadings and Item-Total Correlations of the Team Performance Scale

Item Content	Factor Loading	Item-total Correlation
The services/products delivered by this team are considered satisfactory by the people who receive them.	0.73	0.71
The services of this team are top quality.	0.81	0.78
This team successfully meets its work targets.	0.87	0.84
This team is recognized by top managers for its high performance.	0.75	0.73
This team responds with agility to new demands.	0.81	0.78
The work deadlines set by this team are met.	0.77	0.74
This team is productive.	0.81	0.78
The established targets are met by the team.	0.80	0.77
Other service teams recognize the high performance of this team.	0.70	0.68

Thus, each respondent provided information on their team’s performance. From the operational point of view, the analysis described here was done by calculating the level of similarity of the responses from team members, using the AD_{Md} (interrater agreement index) calculation. A second calculation that confirms the emergence of the construct is the verification of

differences between teams. This analysis was done by calculating the ICC and an Anova (one-way).

The interpretation of the AD_{Md} value is made applying the formula $c/6$ in which “c” represents the amplitude of the scale of responses (Burke & Dunlap, 2002). Since the statements regarding the performance of the team were answered on 5-point scale, the

maximum discrepancy of responses from the members is defined as dividing 5 by 6, resulting in the value of 0.83. The mean value of AD_{Md} for the set of items in the scale was 0.37 (SD = 0.13) with values ranging between 0.09 and 0.67, and therefore being, in all cases, below the value previously mentioned. This demonstrates the existence of similarity (or low discrepancy) in the responses provided by the team members, which is a primary indication that we are looking at an attribute of the teams, despite the data having been collected at the individual level.

The next step was to investigate the existence of differences between teams. The ICC value (0.107) showed that approximately 10.7% of the variance of the phenomenon was derived from level two. In order to interpret the magnitude of the ICC we follow the contributions made by Bliese (2000) who reports, after a review of the empirical work, that the mean value of ICC for studies in the organizational field is 0.12. Therefore, the value found in this study is consistent with others found in the area. Continuing with the task of identify the existence of differences between groups, which certify this is a collective phenomenon, an analysis of variance was done (by ANOVA one-way). The application of this strategy showed that the differences between teams was statistically significant ($F = 1.58$; $DF = 59$; $p \leq 0.01$).

While the measure presents good psychometric properties when used as a tool for self-assessment of team performance, it is worth noting that it was already used by Brito (2014), in a study conducted with teams of airfield firefighters whose performance was evaluated by the supervisor. In this study, the measure was completed by 122 supervisors and the one-factor solution enabled capture of 49.43% of the variance, showing factor loadings that ranged between 0.54 and 0.72, and satisfactory internal reliability ($\alpha = 0.83$ and $r_{\text{item-total}} = 0.55$). In a study done by Reis (2014), the measure was completed by supervisors who evaluated the performance of teams auditors of a federal public institution. In this case, the percentage of variance explained by the one-factor solution was 44.66%, the reliability indices were satisfactory ($\alpha = 0.88$ and $r_{\text{item-total}} = 0.58$), and the factor loadings of the items ranged between 0.45 and 0.70.

The data set obtained in applying the scale to team members, combined with the reports of results from using the measure in assessments made by team supervisors, raises thoughts about its utility, usability and additional applications. The analytical strategies

adopted in the study, which sought to demonstrate the relevance of an instrument designed to assess a meso level attribute, which arises by emergence, should also be discussed. Thus, the following section discusses the theoretical and empirical implications of this research for the field of work team performance, as well as its limitations and possible future developments.

Discussion

Given the wide dissemination of work teams in the world of organizations, there is a natural growing demand for reliable tools to measure the results they achieve. In this regard, Brannick and Prince (2009) point out that the measures available usually focus on the evaluation of concrete results such as the number of takeoffs and landings made without incident, in the case of flight crews, or successful surgical procedures, in the case of medical teams. When the emphasis is on judgmental assessment of results, review of the area literature even shows the marked presence of customized tools, as in the study by McMillan et al. (2013) that describes the process of developing a measure to assess the performance of teams of pilots. These findings support the relevance of the study carried out, focused on the development of a non-specific measure, applicable to different organizational contexts, and that can be answered both by the members themselves as well as by the team supervisor.

In this study, the team performance effectively focused upon was the result (not the processes). To develop the measure, it was also assumed that this result derives from the convergence of individual contributions and is manifested as a collective attribute, which means that it arises through emergence process and is characterized as a meso-level property. When the researcher faces the challenge of elaborating and verifying evidence of the validity of a measure at this level, several precautions should be taken, since the nature of the tool should reflect the theoretical logic of the construct. In the case of studies focusing on work teams it is important to recognize that teams do nothing. Members do the work (Brannick & Prince, 2009). Thus, it is necessary that the measures allow observation of the process of construction and manifestation of a group attribute, yet one that originates at the individual level.

The peculiarity of the phenomena that arise through emergence involves not only the recognition of their theoretical nature, but also requires the adoption of compatible analytical procedures. Regarding their

occurrence, Chan (1998), Klein and Kozlowski (2000), and more recently, Coultas et al. (2014) point out that these are specific phenomena originating from cognitions, motivations, or emotions of individuals (team members), but due to everyday experience of working together, transformations happen and evolve into a collective characteristic. On this new state, they do not allow for distinguishing one member from another, but do differentiate one team from another. This specificity has to be captured by the measure and empirically demonstrated.

These requirements were heeded in conducting this study and thus the theoretical basis that presents and discusses team performance, as well as the analytical strategies chosen, sought to comply with the stipulated guidelines. Thus, the study carried out was based on the understanding that performance is a complex phenomenon that involves processes and results. But, in the study described in this report, the aspect in focus was the results, captured through judgment.

The relevance of adopting the results as performance indicators is supported in the theoretical contributions of the area (Brodbeck, 1996; Hackman, 1987) that defend the importance of acceptance of the product/service by the person who receives it, considered the team's client, whether internal or external. Thus, while performance involves both criteria of processes and results achieved, it is appropriated to focus only on the results that, in this case, were measured by the judgment made by the team members themselves. This decision, although theoretically supported, demands recognition that, with the application of the proposed measure, the manager will hold partial information on the performance of the teams, since the measure does not cover processes.

When the measure is based on judgments made by the members themselves, it should be emphasized that obtaining the individual information is not sufficient to conclude that the measured attribute is actually a characteristic of the team. This is because the individual evaluations may be interpreted under personal perspective of team members, who each see the scenario based on their personal referent. Therefore, in these cases it is necessary to investigate whether the information provided by the team members refer to the members themselves, or in fact to the object (or team) evaluated. In addition, such measures must also be able to discriminate one unit from another (team in this case).

Studies from the organizational psychology field conducted with work teams discuss the relevance of

collecting information from the members, but highlight the need for them to be combined to provide a source for the variables related to the teams (González-Romá & Hernández, 2014; Puente-Palacios et al., 2015; Mathieu et al., 2008; Mohammed, Ferzandi, & Hamilton, 2010). The postulated scale strove to meet this demand, such that the statements of the constituent items focused on the collective, as they inquired about the results achieved by the team. The analysis strategies likewise investigated the similarity of responses given by the team members. Thus the aim was to demonstrate that the target that was effectively evaluated was the team. The need to show reliable evidence for the differentiation between teams was also considered, and for this reason, analyses were performed that revealed its existence.

In addition to the considerations relating to the analytical strategies that confirm the level of the construct, in the process of developing a measure it is necessary to determine the psychometric properties of the instrument. So, a variety of decisions must be made and their consequences estimated.

In the case of the performance evaluation scale, the set of nine items drawn up in the form of descriptors of behaviors, focused on team performance results, proves to be compatible with the theoretical logic of the team performance output variable. Mathieu et al. (2008) refer to empirical studies conducted on the basis of assessments of a judgmental nature, with emphasis on aspects such as overall quality of work, goals achieved, efficiency, and others. Upon analyzing the content covered by the scale items, compatibility is observed with what has already pointed out by studies done in the area. The items with higher loadings, thus those that best represent the construct, refer to achievement of goals, quality of services, and productivity demonstrated.

Even so, it is recognized that since the measure did not derive from interviews or focus groups – which would allow surveying the work team members about what they consider legitimate indicators of performance –, this may have excluded important aspects. However, since this study has relied on the participation of expert judges from the area, who evaluated the constituent items of the measure and could propose others, it is considered that the phenomenon of interest can be satisfactorily captured by the proposed set of items.

The factor solution retained allows for the capture, in self-assessment situations like those in this

study – when the team members themselves evaluated the performance of their work unit – of a high percentage of the variance of the phenomenon (65.84%), which can be interpreted as evidence of the relevance of the retained items' content and the analytical decisions made.

The strategy adopted to verify the factor structure of the measure consists, as Laros (2005) indicates, of a set of choices made by the researcher, in order to verify to what extent the data matrix can be appropriately reduced to factors, capturing the greatest amount of variance and, in a concomitant manner, favoring parsimony by grouping items into factors. In this sense, the decisions on which this study was guided prove appropriate both by the percentage of variance captured and by the content of the items that show adherence to the theoretical bases of the construct (Mathieu et al., 2008; González-Romá & Hernández, 2014).

The evidence found in the process of reducing the data matrix to a single factor show the relevance of the decisions that were made, because there are no theoretical indications that suggest the need to retain a greater number of factors when the target of the assessment is the results or output criteria of the team's work. So, from the results concerning its psychometric properties, it can be justly stated that the measure provides a reliable assessment of the performance of the teams, according to the data obtained in the sample that was investigated.

In addition to these attributes, the results obtained from analyzing the patterns of variance should be discussed. In this regard, Puente-Palacios and Borba (2009) caution about the need to identify similarity within groups and differences between them, as a minimum requirement for verifying that data collected from the individuals adequately represent the group. The research reported here identified low discrepancy between the responses of the members and significant variance between the teaching teams, which demonstrates that the measure is an instrument capable of capturing a group property.

Specifically regarding the fact that the individual responses allow legitimate derivation of constructs at the meso level, Puente-Palacios and Martins (2013) argue the need for alignment between the level to which the phenomenon theoretically pertains and the level at which the analyses are carried out. These authors also point out that the data can be collected on a different level, but emphasize that analytical strategies will have to be adopted so as to compose scores that genuinely

represent the attribute in question. From this it is concluded that the measure elaborated here captures the group performance, even when the answers are collected from the team members.

As to the usefulness of the proposed scale, it is important to stress that its use is not restricted to situations of self-evaluation done by the team members themselves. It can also be used as an instrument for assessment done by a manager or supervisor. Thus, it serves as a tool that can be used by multiple sources, as the psychometric properties that it presents in these situations are also quite favorable (Brito, 2014; Reis, 2014).

Despite the contributions of this study, some limitations can be noted, such as the fact that a sample of workers from a single sector was used. Thus, the evidence reported on psychometric validity was obtained from a single sample and may reflect the specific nature of this group of respondents. Another limitation was the fact that the sample forming the basis of this study came from another Latin American country; and so the measure should be used with caution in national studies. Even so, subsequent applications conducted with the Portuguese version of the measure (Brito, 2014; Reis, 2014) have shown promising results.

The need to investigate the predictive capacity of the measure, as compared to other performance criteria, should also be pointed out. It would therefore be right to expect significant positive associations between the performance assessment diagnosed by the proposed measure and that done using scales focusing on group processes, such as cohesion, coordination, and communication, given that group processes are also considered performance indicators (Beal et al., 2003; Brannick and Prince (2009) Further studies using the measure for assessing work team performance should also be conducted to find evidence of discriminant validity, for example, with an individual performance evaluation measure. That would make it possible to obtain empirical indications that the scale elaborated here actually focuses on a collective phenomenon, in this case the team's performance, and differs from measures that focus on individual performance.

The theoretical route taken over the course of this research, combined with empirical results obtained in the testing of the measure, allow us to derive some practical implications. Among them it is worth considering the fact that in Brazil there is no collective assessment measure of team performance. Thus, the instrument developed here constitutes a performance diagnostic

tool for units increasingly present in the organizational setting. Additionally, it is important to note that the measure focuses on a meso level attribute, a fact that constitutes a distinction among organizational behavior studies, in which the preferred focus tends to be the individual. Third, the manuscript describes the methodological course to be followed in the development of a measure that seeks to assess a collective attribute that arises by emergence. The importance of this contribution stems from the fact that the teams and their members experience various processes, many of which can be captured by adopting strategies similar to those reported in this study. These proposals, however, must be accompanied by empirical verification because, as demonstrated, it is the theoretical nature of the attribute that determines the characteristics that the measure presents. Thus, further research needs to be conducted in order to contribute to the advancement of knowledge in this field.

References

- Beal, D., Cohen, R., Burke, M., & McLendon, C. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology, 88*(6), 989-1004.
- Bell, S. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology, 92*, 595- 615.
- Bliese, P. D. (2000). Within-group agreement, non-independence and reliability. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions and new directions* (pp. 349-381). San Francisco, CA: Jossey-Bass.
- Brannick, M., & Prince, C. (2009). An overview of team performance measurement. In M. Brannick, E. Salas & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods and applications* (pp. 3-16). New Jersey: Psychology Press.
- Brito, L. (2014). *Ser competente basta? Um estudo sobre competências necessárias para o trabalho em equipes de bombeiros de aeródromos* (Dissertação de mestrado). Universidade de Brasília, Brasília, DF.
- Brodbeck, F. C. (1996). Criteria for study of work group functioning. In M. A. West (Ed.), *Handbook of work group psychology* (pp. 285-315). San Francisco: Jossey-Bass.
- Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: A user's guide. *Organizational Research Methods, 5*(2), 159-172.
- Callea, A., Urbini, F., Benevene, P., Cortini, M., Lemma, O. L., & West, M. (2014). Psychometric properties and factor structure of the Italian version of the "Aston team performance inventory". *Team performance management, 20*, 6-18.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*, 234-246.
- Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23*, 239-290.
- Coultas, Ch., Driskell, T., Burke, C. S., & Salas, E. (2014). A conceptual review of emergent state measurement: Current problems, future solutions. *Small Group Research, 45*, 671-703.
- DeNisi, A. (2000). Performance appraisal and performance management. A multilevel analysis. In S. Kozlowski & K. Klein (Eds.), *Multilevel theory, research and methods in organizations*. (pp. 120-156). San Francisco: Jossey-Bass.
- DeShon, R., Kozlowski, S., Schmidt, A., Milner, K., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89*, 1035-1056.
- González-Romá, V., & Hernández, A. (2014). Climate uniformity: It's influence on team communication quality, task conflict, and team performance. *Journal of Applied Psychology, 99*, 1042-1058.
- González-Romá, V., Fortes-Ferreira, L., & Peiró, J. M. (2009). Team climate, climate strength and team performance. A longitudinal study. *Journal of Occupational and Organizational Psychology, 82*, 511-536.
- Hackman, J. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of Organizational Behavior* (pp. 315-342). New York: Prentice-Hall.
- Klein, K., & Kozlowski, S. (2000). *Multilevel theory, research and methods in organizations*. San Francisco: Jossey-Bass.

- Kozlowski, S., Chao, G., Grand, J., Braun, M., & Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods*, 16(4), 581-615.
- Laros, J. A. (2005). O uso da análise fatorial: Algumas diretrizes para pesquisadores. In L. Pasquali (Eds.), *Análise fatorial para pesquisadores* (pp. 163-184). Brasília: LabPAM.
- Mathieu, J. E., Maynard, M. T., Rapp, T., & Gilson. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, 34, 410-476.
- McMillan, J., Entin, E., Morley, R., & Bennett Jr. W. (2013). Measuring team performance in complex and dynamic military environments: The SPOT-LITE Method. *Military Psychology*, 25, 266-279.
- Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publications.
- Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, 36, 876-910.
- Nadler, D., Hackman, R., & Lawler, E. (1979). *Managing organizational behavior*. Boston: Little, Brown, & Company.
- Puente-Palacios K. E., & Martins, M. C. F. (2013). Gestão do clima organizacional. In L. Borges & L. Mourão (Eds.), *O trabalho e as Organizações. Atuações a partir da Psicologia* (pp. 251-278). Porto Alegre: Artmed.
- Puente-Palacios, K. E., & González-Romá, V. (2013). Gestão de equipes. In L. Borges & L. Mourão. *O trabalho e as Organizações. Atuações a partir da Psicologia*. (pp.311-338). Porto Alegre: Artmed.
- Puente-Palacios, K. E., & Borba, A. C. (2009). Equipes de trabalho: Fundamentos teóricos e metodológicos da mensuração de seus atributos. *Avaliação Psicológica*, 8(3), 369-379.
- Puente-Palacios, K., Silva, R. da, & Borba, A. C. (2015). Potência em equipes de trabalho. In K. Puente-Palacios & Peixoto, A. *Ferramentas de diagnóstico para organizações e trabalho: um olhar a partir da psicologia*. (pp. 187-205). Porto Alegre: Artmed.
- Priesemuth, M., Schminke, M., Ambrose, M., & Folger, R. (2014). Abusive supervision climate: A multiple-mediation model of its impacts on group outcomes. *Academy of Management Journal*, 57, 1513-1534.
- Quigley, N., Tekleab, A., & Tesluk, P. (2007). Comparing consensus- and aggregation-based methods of measuring team-level variables the role of relationship conflict and conflict management processes. *Organizational Research Methods*. 10(4), 589-608.
- Reis, D. (2014). *O papel da identidade social na efetividade de equipes de trabalho* (Dissertação de mestrado). Universidade de Brasília, Brasília, DF.
- Salas, E., Cooke, N., & Rosen, M. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50, 540-547.
- Smith-Crowe, K., J., Burke, Kouchaki, M., & Signal, S. (2013). Assessing Interrater agreement via the average deviation index given a variety of theoretical and methodological problems. *Organizational Research Methods* 16(1), 127-151.
- Sonnentag, S. (2002). *Psychological management of individual performance*. Chichester: John Wiley & Sons, LTD.

Recebido 28/03/2015

1ª Reformulação 05/08/2015

2ª Reformulação 09/09/2015

Aceito 21/09/2015

Sobre os autores:

Katia Puente-Palacios é mestre e doutora em Psicologia pela Universidade de Brasília (UnB) e atua como docente no Programa de Pós-graduação em Psicologia Social, do Trabalho e das Organizações, na Universidade de Brasília. Seu campo de interesse em ensino e pesquisa é relativo ao funcionamento de equipes de trabalho.

E-mail: puente@unb.br

Maria do Carmo Fernandes Martins é mestre, doutora em Psicologia pela Universidade de Brasília e professora do Programa de Pós-graduação em Psicologia da Saúde da Universidade Metodista de São Paulo. Atua na área de Comportamento Organizacional, Psicologia Organizacional e do Trabalho, Medidas em Psicologia e Psicologia da Saúde.

E-mail: mcf.martins@uol.com.br

Salvatore Palumbo é pós-graduado em Gestão do Conhecimento e Inteligência Empresarial pela Pontifícia Universidade Católica do Paraná (PUC-PR) e em Gestão Estratégica de Pessoas pela Fundação Instituto de Administração (FIA), auditor do Tribunal de Contas da União (TCU) e professor do Instituto Serzedelo Corrêa.

E-mail: Salvatore@tcu.gov.br

Contato com os autores:

Universidade de Brasília - Campus Darcy Ribeiro
ICC Sul. Instituto de Psicologia, Sala AT-013