Papeles del Psicólogo, 2014. Vol. 35(2), pp. 115-121
http://www.papelesdelpsicologo.es

RELEVANT FACTORS TO INCREASE THE ACCURACY, FEASIBILITY AND SUCCESS OF JOB PERFORMANCE EVALUATION SYSTEMS

Dolores Díaz Cabrera, Estefanía Hernández Fernaud, Rosa Isla Díaz, Naira Delgado Rodríguez, Luis Díaz Vilela and Christian Rosales Sánchez
Universidad de La Laguna

The situation of economic crisis in Spain has motivated an increasing interest in employee performance evaluation in organizations. On the other hand, it is also producing greater resistance to this evaluation on the part of appraisees, who regard these processes with suspicion and concern in these times of economic difficulties. Therefore, in economic situations like the current one, it is very important to develop assessment scales that facilitate the effectiveness, credibility and success of these systems. The general goal of this work is to provide results that can help evaluation systems to be viewed positively by organizations as well as by their members. For this it is necessary that: (a) the response scales used offer precise and reliable data; and (b) the procedures applied to assess performance are perceived as fair and reliable by the participants.
Key words: *Work performance assessment/appraisal, Task and citizenship performance scales, Reactions to performance assessment/appraisal, Perceived fairness/justice.*

La situación de crisis económica en nuestro país ha provocado, por una parte, un incremento del interés de las organizaciones en la evaluación del desempeño laboral. Por otra parte, está produciendo una mayor resistencia a la evaluación por parte de los evaluados, que miran con desconfianza y preocupación su evaluación en momentos de dificultades económicas. Por ello, en situaciones económicas como la actual, es muy importante desarrollar escalas de evaluación que permitan lograr la efectividad, la credibilidad y el éxito de estos sistemas. El objetivo general de este trabajo es aportar resultados dirigidos a que los sistemas de evaluación sean valorados positivamente tanto por la organización como por los participantes. Para ello es necesario que: (a) las escalas de respuesta usadas ofrezcan datos precisos y fiables; y (b) los procedimientos aplicados para la evaluación sean percibidos como justos y fiables por los participantes.
Palabras clave: *Evaluación del desempeño laboral, Escalas de desempeño de tareas y cívico, Reacciones a la evaluación del desempeño, Justicia percibida.*

The economic crisis in our country has increased the interest of organizations in the study of efficiency and job performance. However, it has also increased the resistance to evaluation by appraisees and even evaluators, who regard appraisals with suspicion and concern in these difficult economic times. This resistance is mainly the result of the fact that performance evaluation is usually associated with negative consequences for workers. And these consequences are perceived even more negatively in situations of economic crisis.

Resistance to evaluations, and the reactions associated with them by workers, quite often cause organizations to delay the implementation of these evaluation systems, being immersed in reorganization, restructuring or simply due to the necessary caution with which they regard the question of evaluating staff performance in these times.

However, other organizations in the current economic situation regard the improvement of job performance as one of their central objectives in order to increase their effectiveness. A major risk that arises from this scenario is that the urge to produce performance indicators as quickly as possible hinders the transparency and the accuracy of the evaluation, and therefore, reduces the confidence of those involved in these processes.

In economic crises such as the current one, it is even more important to consider the three criteria proposed by Cardy and Dobbins (1994), which aim to ensure the effectiveness, credibility and success of these systems. These three criteria are: (a) to consider the qualitative aspects of the evaluation process; (b) to reduce the biases of the performance evaluators themselves; and (c) to increase the accuracy of the measures used for evaluation.

This article presents the results obtained by our research team in several studies that address issues related to these three criteria. Funded by an R&D agreement of the Canary Islands'

Correspondence: Dolores Díaz Cabrera. *Departamento de Psicología Cognitiva, Social y Organizacional. Facultad de Psicología. Campus de Guajara. 38205 La Laguna. Tenerife. España. E-mail: mddiaz@ull.es*

....

This research has been developed with the joint funding from the agreement Designing a performance evaluation system for public employees of the Canary Islands Government, and the National Research Plan project P512010-17327

Public Administration 2009-2010, a performance appraisal system for public employees of the Public Administration of the Autonomous Region was designed. Also, our team is currently carrying out a research project funded by the National Plan (2011-2014) on the relationship between job performance and some personal and organizational variables, and the role of training for evaluators in performance appraisal.

This article deals with various factors that we believe are relevant to evaluation systems being viewed positively both by the organization and by appraisees and evaluators, thereby reducing the resistance to participation. In our view, this requires, firstly, that the evaluation procedures are perceived as fair by the participants, which will be achieved through, amongst other things, providing information on the evaluation system and encouraging the participation of those involved. Secondly, the response scales used must provide accurate and reliable data, and this depends on the first two criteria presented above being taken into consideration: the reduction of bias and the development of accurate scales.

Specifically, this article explores, first, the variables that affect the perceptions of workers regarding the implementation of a performance appraisal system (e.g. attitudes and beliefs, the perceived justice and the acceptance of these systems). Second, we analyze the psychometric properties of two rating scales (the Likert-type scale -or graphical behavior scale- and the standardized mixed type scale), and their effects on response bias. Before addressing these points, in the following section we discuss some basic concepts regarding job performance and its relevance to professional practice.

THE IMPORTANCE OF PERFORMANCE EVALUATION IN HUMAN RESOURCE MANAGEMENT

Performance appraisal has important advantages for both organizations and workers. One of its main contributions is to increase the organization's awareness of how the work is being performed. The information obtained will also guide the design and development of the selection processes and the detection of training needs, including human resource management strategies. Also, other uses of evaluations for organizations are related to the recognition of workers and the development of motivational plans. Meanwhile, workers will receive information that helps them to have greater clarity regarding their roles and tasks, to become aware of how they carry out their work and to improve their work practices.

Performance evaluation has been the subject of interest of many researchers and practitioners. A general definition of job performance comprises the observable and measurable behaviors and results that involve employees or they participate in, contributing to the achievement of organizational goals (Viswesvaran & Ones, 2000).

In the performance evaluation it is considered important to include all of the behaviors that impact positively or negatively on the organizational results, and not just the tasks formally assigned to the task and specific position (Salgado & Cabal, 2011). Since the nineties there has been a distinction made between two types of job performance, task performance and contextual performance (also known as citizenship performance). The first concerns the role requirements to be met by the employee, directly related to the efficacy with which he or she performs the activities that contribute to organizational goals (Borman & Motowidlo, 1993).

Citizenship performance is defined as the set of behaviors that are not directly related to the execution of the tasks assigned to the job, but which have great importance, since they promote an appropriate organizational, social and psychological context for the effective accomplishment of the given tasks (Borman & Motowidlo, 1993). What is important in this type of performance is the initiative, the support and the persistence that the employee shows rather than the technical competence demonstrated (Poropat, 2002).

In the following sections we address some of the results obtained by our research team in three studies that have had as a general objective to achieve greater effectiveness and success of the performance evaluation systems and procedures. Specifically, in the following section we will focus on one of the three criteria outlined by Cardy and Dobbins (1994): the qualitative aspects of the evaluation system. The main proposals that have been made regarding these aspects are briefly described, as well as some of the results obtained in our research.

PROMOTING PARTICIPANTS' ACCEPTANCE OF THE EVALUATION SYSTEM: PERCEIVED JUSTICE AND TRUST

Most studies on performance evaluation have focused mainly on the development of response scales that reduce, to the greatest extent possible, the biases in evaluation, mainly benevolence bias or the tendency of evaluators to assign high scores. Although these objectives are very important in achieving accurate assessments, there has been criticism of the lack of focus on other criteria such as the participants' reactions to performance appraisals and the factors contributing to these reactions (Cardy & Dobbins, 1994; Murphy & Cleveland, 1995, cf. Cawley, Keeping & Levy, 1998). This lack of attention has led Murphy and Cleveland (1995) to refer to this third category, the perceptions and reactions of the participants, as "neglected criteria" (p. 310).

Workers' attitudes towards evaluation systems influence their participation in performance evaluations. In a survey conducted in the UK (Industrial Society, 1997, cf. Fletcher, 2002) on participation in performance evaluation processes, it was obtained that less than two-thirds of the organizations rated achieved 67% of all possible evaluations. Indeed, even when there is a system of highly accurate evaluation, if it is not accepted and supported by the members of the organization, its effectiveness and feasibility will be limited (Farr & Jacobs, 2006; Keeping & Levy, 2000).

The importance of the social and organizational context in which evaluations are carried out is emphasized in what Folger, Konovsky and Cropanzano (1992) refer to as the metaphor of "due-process". This metaphor defends the importance of developing processes taking into account aspects such as the following: (a) participants must have prior information about the objectives and criteria that will be used and they may be involved in the development of these, or at least express their opinion of them; (b) appraisees must have the opportunity to explain the results obtained in the evaluation; (c) evaluators must be familiar with the jobs to be evaluated and the evaluation system; and (d) evaluators must have the opportunity to make evidence-based judgments, without external pressures or personal bias, and based on values such as honesty and justice.

One of the constructs with the greatest impact on the acceptability and effectiveness of an evaluation system is the one that Farr and Jacobs (2006) propose to call "the collective confidence of the participants in the evaluation process" (p. 324), which reflects the idea that participants must trust the process before the evaluation can be effectively implemented in the organization.

For its part, organizational justice theory emphasizes the importance of similar constructs, emphasizing that the success of the evaluation process will depend on the reactions of the participants. These reactions depend, amongst other things, on: (a) the attitudes and beliefs of the participants on issues such as degree of comfort with the questionnaires and rating scales used, their beliefs about the purpose of these evaluations, satisfaction with the evaluator, the evaluators' perceptions of their own ability to evaluate, etc.; (b) the evaluation context: for example, the degree of participation of members of the organization in the development of the evaluation process and their satisfaction with the evaluation system; and (c) satisfaction with the feedback of the results of the evaluation (Giles & Mossholder, 2000; Levy & Williams, 2004; Tziner, Murphy & Cleveland, 2001).

Other studies have also shown the relationship between variables such as affective commitment to the organization (positive feelings of identification, involvement and attachment to the organization) and the perception of the evaluation system as fair (Colquitt, Conlon, Wesson, Porter & Yee, 2001; Colquitt et al, 2013; Dirks & Ferrin, 2001; Salleh, Amin, Muda & Halim, 2013). Therefore, the acceptance of evaluation systems by the participants and their degree of involvement in them is determined by their trust in the specific process developed and by their perceived justice in relation to its development and implementation and the feedback that it provides.

The study conducted by our research team aimed to explore the impact of some of the variables discussed above on the perceptions of employees towards the implementation of a performance evaluation system that was being designed at the time and which was to be implemented in the short term, and for the first time, in an organizational unit (Díaz-Cabrera, Hernández-Fernaud, Rosales-Sánchez, Isla & Díaz-Vilela, under review).

Specifically, we analyzed the extent to which some attitudes and beliefs (Job satisfaction, Attitude to change, Self-efficacy, Initiative and Affective commitment) explain the Perceived justice of the evaluation, and the Acceptance of the future evaluation system. One hundred and forty-two civil servants participated, constituting the total workforce in one unit of a regional administration.

The results obtained indicate that the social and organizational context in which the evaluation is carried out together with the reactions of participants explain, at least partially, the employees' perceptions of the performance evaluation system to be implemented. Specifically, in the first place, the results show that workers with a higher affective commitment to the organization tend to evaluate more highly the implementation of a system for evaluating their performance. And secondly, the acceptance of the performance evaluation system by the participants was explained at 13.4% by job satisfaction, affective commitment and attitude towards the changes that may occur in the employment context.

These results highlight the importance of considering participants' reactions to evaluation systems in order to facilitate their implementation. In the following section we address another important aspect in achieving the effectiveness and acceptance of these systems.

PROMOTING THE REDUCTION OF EVALUATION BIAS THROUGH THE PRECISION OF MEASUREMENT SCALES

Folger et al. (1992) use the "test metaphor" to describe the implicit assumptions underlying the development of rating scales. These assumptions are as follows: (a) the nature of the job must allow valid and reliable measures of behavior and work performance; (b) evaluators must be able to and want to make precise assessments; and (c) there must be a true and rational criteria evaluation for each position.

The criteria proposed by Cardy and Dobbins (1994) aim to improve on these assumptions. Specifically, of the three aforementioned approaches for improving the success of the evaluation systems, two of them directly affect the development of the scales, favoring the reduction of the biases that evaluators may have in appraising performance and increasing the accuracy of the measurement instruments.

Although there are different types of evaluation bias (e.g., halo, severity, primacy), benevolence is one of the most frequent. Benevolent bias is the tendency for evaluators to assign higher scores in evaluations (Aguinis, 2013; Guilford, 1954). This bias has a significant impact on the performance appraisal process as it reduces the possibility of correctly recognizing and rewarding employee performance, and therefore the validity of evaluations decreases considerably (Bretz, Milkovich & Read, 1992). The occurrence of this bias was initially attributed to situational and organizational factors, with evaluator variables being considered later as a possible cause. Thus, the recurrent positive evaluation of performance has been related to the purpose of the performance evaluation, the avoidance of confrontation with appraisees, the existence of personal relationships between the evaluators and appraisees, and certain personality characteristics of the evaluator, among other variables (Bernardin, Cooke & Villanova, 2000).

Reducing benevolent bias has traditionally focused on the development of training programs for evaluators, as well as the development of improved formats for response scales (Woehr & Huffcutt, 1994). Previous research has shown that the design of the measurement scale influences the distribution of responses on the scale. Specifically, Viswesvaran and Ones (2005) argue that forced-choice scales and mixed standard scales (MSS) can reduce the benevolent bias when evaluators intentionally try to bias their responses. Our research team has worked in recent years on the development of a response scale format that is different from the traditional graphic scale, comparing the effect that each one has on benevolent bias in evaluating task performance. With these objectives in mind, we have carried out two investigations, one with administrative workers, and other with university students.

In the first investigation, firstly, a job post analysis was carried out using questionnaires and interviews in order to identify the most relevant administrative tasks in each job and to enable performance evaluation. From this information an inventory of tasks to be carried out by the workers was developed, assessing the frequency, importance and complexity of each task in the job. A committee of experts made up of researchers, supervisors and job occupants, evaluated each of the administrative tasks according to their relevance, including a total of 60 tasks (Díaz-Vilela, Delgado, Isla-Díaz, Díaz-Cabrera & Hernández-Fernaud, under review).

With these 60 administrative tasks a scale of observed behaviors (Behavior Observation Scale BOS) was developed for performance evaluation, which was completed in pencil and paper format, by 53 employees of a public administration in administrative positions. The workers had to assess the frequency of performing each of the tasks properly, using a graphical ten-point scale. It was found that 14 of the tasks included in the questionnaire were performed by most administrators, while the remaining tasks had greater variability among the jobs (Díaz-Vilela et al., 2013).

Secondly, a new tool was developed for evaluating the performance of workers performing administrative functions, using a mixed standardized scale format (MSS). This instrument included the same 14 previous tasks, but the wording and presentation of the statements was different. This new scale was presented using a computer app. For each task, three response options appeared on the screen representing different levels of performance (Excellent, Good, Needs improvement). These response alternatives were presented in random order and position in each of the tasks evaluated. That is, the responses were not always ordered from poor performance to excellent or vice versa. After selecting the level of performance, other response alternatives opened in which users had to indicate how often that level of performance is carried out.

With this new scale of response the aim is: 1) to avoid the social desirability of responses, with performance levels not being labeled at as Excellent, Good or Needs improvement, and not necessarily in order from the worst to the best performance, or vice versa; 2) to encourage evaluators to pay more attention to the content of the answer options, because as they appear in a different order and position each time they require a conscious reading of each option to select the desired level of performance; and 3) to reduce benevolent bias, decreasing the possibility of responding to each item automatically and requesting an assessment of the frequency with which the task is performed at the selected performance level (Sometimes, Usually or Always).

In a first study comparing these two response scale formats, the Likert-type graphic scale and the MSS, the data obtained in two independent samples of administrators were compared (Díaz-Cabrera et al., 2013). Of the 61 administrators included in the sample, 37.7% self-evaluated their performance and were rated by their supervisors with the BOS scale with Likert-type graphic scale in pen and paper format, and the remaining 62.3% evaluated their performance and were assessed through the computer app with the questionnaire using an MSS response scale. Thus, each participant had two measures: self-evaluation of their performance, and the evaluation carried out by their supervisor.

The results of this study showed a lower tendency to assign higher scores on performance assessment when the MSS response scale was used in comparison with the traditional graphic scale. This means that benevolent bias was reduced. Supervisors, in comparison with self-assessments, assign lower scores, both when a Likert-type scale is used and when the evaluation is performed with an MSS scale.

Furthermore, a higher standard deviation was observed using the MSS scale in comparison with the Likert-type scale. That is, those who used the MSS scale assigned different scores more frequently depending on the particular task being evaluated, reducing the tendency of evaluators to assess the worker's performance similarly on each of the tasks. Supervisors also showed greater variability of response in their performance evaluations when using the MSS scale rather than the graphic scale.

These data support our hypothesis that using the MSS scale reduces benevolent bias and favors greater intra-subject variability of responses than when using a Likert-type scale. However, this first study has two important limitations from our point of view: the small size of the sample, and the fact that the samples that evaluated the performance with two types of scale were different, i.e., an inter-subject design was used. In order to address these shortcomings and to confirm the results obtained the second investigation was conducted.

The objective of this second study was to analyze whether significant differences exist in the performance evaluation of the same person when using a scale of observed behaviors with Likert-type response scale, as opposed to a mixed standardized type scale with behavioral anchors (Hernández-Fernaud et al., 2013). As in the previous investigation, two instruments were developed comprising the same behaviors to be evaluated, eight tasks necessary for group work, and they varied only in the type of response scale.

The first phase of this research was conducted with 68 university students who had carried out a practical activity in a small group, which was the object of the performance evaluation. A repeated measures design was used, such that each participant self-evaluated their performance in the group work twice, once using the graphic scale, and the other using the MSS scale. The order of completion of the scales was also considered as a variable, so half of the sample answered using the graphic scale first and a month later the MSS, and the other half of the sample answered using the scales in the reverse order. Both scales were presented using a computer app. The results showed a statistically significant effect of the type of scale, such that when the MSS scale was used the self-assessment of performance was lower than when a Likert-type graphic scale was used. It was also observed that the order of completion of the scales did not change the results.

In the second phase of this research the sample was extended to 188 participants, without taking into consideration the order of completion of the two instruments. The results were in line with the previous ones; the average self-assessment of performance was significantly higher using the Likert-type scale than using the MSS scale, therefore the latter reduces the benevolent bias. Also, as in the first investigation discussed, it was observed that the average standard deviation obtained with the graphic scale was lower than that obtained for the MSS range, such that participants scored their answers in a wider range with the MSS than when using the graphic scale.

The results that we have discussed indicate that using instruments with response scales presenting behavioral anchors without an explicit identification of the performance level they represent decreases the tendency to evaluate performance positively, and favors the evaluator to use more points on the scale of response than when a traditional graphic scale is used, reducing the benevolent bias in job performance evaluations.

CONCLUSIONS AND PRACTICAL IMPLICATIONS

Performance evaluation is a vital strategy in people management in organizations and it is particularly relevant in situations such as the current economic crisis. Performance evaluation plays a dual role in organizations. On the one hand, it contributes to the improvement of organizations as it facilitates the achievement of the organizational goals of efficiency, effectiveness and transparency, because, among other things, it enables the redesigning of jobs, the development of training plans and selection processes and decision-making regarding promotions and incentives policies. On the other hand, it affects the motivation of workers through involving them in the process and through the feedback received on their performance, allowing them to improve and be recognized and compensated.

To achieve these positive effects of performance evaluation, it is important to keep in mind two objectives. As Farr and Jacobs (2006) suggested, it is necessary for the organization and its participants to view performance evaluations positively, which will depend on a series of conditions. Specifically, the organization will consider positively having accurate and reliable performance data, which will depend on the use of precise evaluation scales and a reduction in the biases of the evaluators (objectives of the test metaphor). Furthermore, participants will view the evaluations more positively if they are fair and credible, and this depends, in turn, on the accuracy of the evaluation (test and due-process metaphors). In addition, the perceived justice, and therefore, the success and effectiveness of the evaluation system, will depend on: (a) the procedure used recognizing the different individual goals and motivations at stake; (b) the system to be used providing clear information to participants; and (c) the evaluators and appraisees participating in the process of designing the system.

REFERENCES

Aguinis, H. (2013). *Performance management*. Upper Saddle River, New Jersey: Pearson

Bernardin, H.J., Cooke, D.K., & Villanova, P. (2000). Conscientiousness and Agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, *85*(2), 232-234. http://dx.doi.org/10.1037//0021-9010.85.2.232

Borman, W. C., & Motowidlo, S. (1993). Expanding the criterion domain to include elements of contextual performance. in N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations*, (pp. 71-98). San Francisco: Jossey - Bass.

Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, *18*, 321-352. http://dx.doi.org/10.1177/014920639201800206

Cardy, R. L., & Dobbins G. H. (1994). *Performance appraisal: alternative perspectives*. South-Western, Cincinnati: OH.

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the Performance Appraisal Process and Employee Reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, *83*(4), 615-633. http://dx.doi.org/10.1037//0021-9010.83.4.615

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O.L. H., & Yee Ng, K. (2001). Justice at the Millennium: A Meta-Analytic Review of 25 years of Organizational Justice Research. *Journal of Applied Psychology*, *86*(3), 425-445. http://dx.doi.org/10.1037/0021-9010.86.3.425

Colquitt, J. A., Scott, B. A., Rodell, J. B., Long, D. M., Zapata, C. P., Colon, D. E., & Wesson, M. J. (2013). Justice at the Millennium, a Decade Later: A Meta-Analytic Test of Social Exchange and Affect-Based Perspectives. *Journal of Applied Psychology*, *98*(2), 199-236. http://dx.doi.org/10.1037/a0031757

Díaz-Cabrera, D., Díaz-Vilela, L., Delgado, N., Isla-Díaz, R., Rosales-Sánchez, C. & Hernández-Fernaud, E. (2013). *Comportamiento de las distribuciones de medidas del desempeño en función del tipo de escala utilizada. [Behavior of distributions of performance measures depending on the type of scale used]*. Paper presented at the XIII Congress of Methodology of Social Sciences and Health (AEMCCO), La Laguna, Spain.

Díaz-Cabrera, D., Hernández-Fernaud, E., Rosales-Sánchez, C., Isla, R. & Díaz-Vilela, L. (under review). Aceptación y justicia percibida de la evaluación del desempeño: análisis de variables predictoras [Acceptance and perceived fairness of performance evaluation: analysis of predictor variables].

Díaz-Vilela, L., Delgado, N., Isla-Díaz, R., Díaz-Cabrera, D. & Hernández-Fernaud, E. (under review). Relationships between contextual and task performance and interrater agreement: Why there are none?

Díaz-Vilela, L., Díaz-Cabrera, D., Isla-Díaz, R., Hernández-Fernaud, E., Rosales-Sánchez, C. & Delgado, N. (2013). *Development of a computer based task performance scale for clerical workers in a Public Administration*. 16th Congress of the European Association of Work and Organizational Psychology (EAWOP), Münster, Germany.

Dirks, K. T. & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science*, *12*, 450-467. http://dx.doi.org/10.1287/orsc.12.4.450.10640

Farr, J. L & Jacobs, R. (2006). Trust us: New perspectives on Performance appraisal. In W. Bennett, C. Lance & D. Woehr (Eds.), *Performance Measurement: Current perspectives and Future challenges* (pp. 321-337). Mahwah, New Jersey: LEA, Publishers.

Fletcher, C. (2002). Appraisal: An individual psychological perspective. In S. Sonnentag (Ed.), *Psychological management of individual performance* (pp. 115-136). New York, NY: John Wiley & Sons, Ltd.

Folger, R., Konovsky, M., & Cropanzano, R. (1992). A due process metaphor for performance appraisal. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior*, Vol. 14 (pp. 127-148). Greenwich: JAI Press.

Giles, W. F., & Mossholder, K. W. (1990). Employee reactions to contextual and session components of performance appraisal. *Journal of Applied Psychology*, *75*, 371-377. http://dx.doi.org/10.1037//0021-9010.75.4.371

Guilford, J. P. (1954). *Psychometric method*. Nueva York: McGraw-Hill.

Hernández Fernaud, E., Díaz-Vilela, L., Isla-Díaz, R., Delgado, N., Rosales-Sánchez, C. & Díaz-Cabrera, D. (2013). *Efecto del tipo de escala de medida en la evaluación del desempeño en estudiantes universitarios [Effect of the type of measurement scale on performance evaluation in university students]* Paper presented at the XIII Congress of Methodology of Social Sciences and Health (AEMCCO). La Laguna, Spain.

Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: measurement, modeling, and method bias. *Journal of Applied Psychology*, *85*, 708 - 723. http://dx.doi.org/10.1037//0021-9010.85.5.708

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*(6), 881-905. doi:10.1016/j.jm.2004.06.005

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational and good-based perspectives*. Thousand Oaks: CA: Sage.

Poropat, A. E. (2002). New Models of Work Performance and Their Implications for Employment Relations. In P. Ross, M. Lyons, C. Allan & K. Townsend (Eds.), *Proceedings of the Tenth Annual Conference of the International Employment Relations Association*. Sydney: International Employment Relations Association (IERA).

Salgado, J. F. & Cabal, A. L. (2011). Evaluación del desempeño en la Administración Pública del Principado de Asturias: Análisis de las propiedades psicométricas [Performance Evaluation of Public Administration in the Principality of Asturias: Analysis of the psychometric properties]. *Revista de Psicología del Trabajo y de las Organizaciones*, *27*(2), 75-91.

Salleh, M., Amin, A., Muda, S., & Halim, M. A. S. A. (2013). Fairness of performance appraisal and organizational commitment. *Asian Social Science*, *9*(2), 121-128. http://dx.doi.org/10.5539/ass.v9n2p121

Smith, P. (1976). Behavior, results, and organizational effectiveness: The problem of criteria. In M. D. Dunnette (Ed.), *Handbook of Industrial and organizational psychology* (pp. 745-775). Chicago: Rand McNally College Pub. Co.

Tziner, A., Murphy, K. R., & Cleveland, J. N. (2001). Relationships between attitudes toward organizations and performance appraisal systems and rating behavior. *International Journal of Selection and Assessment*, *9*(3), 226-239. http://dx.doi.org/10.1111/1468-2389.00176

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*(4), 216-226. doi:http://dx.doi.org/10.1111/1468- 2389.00151

Viswesvaran, C., & Ones, D. S. (2005). Job performance: Assessment issues in personnel selection. In A. Evers, N. Anderson & O. Voskuill (Eds.), *The Blackwell Handbook of Personnel Selection*, (pp. 354-375). Malden: Blackwell Publishing.

Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, *67*, 189-205.