Empirical Research on Multisource Feedback in Performance Appraisal: A Literature Review

Mariana Lobo Moreno[1], Nuno Rebelo dos Santos[2],
Leonor Maria Pais[3]

[1] http://orcid.org/0000-0002-5842-184X / Universidade de Coimbra, Portugal
[2] http://orcid.org/0000-0002-2087-5950 / Universidade de Évora, Portugal
[3] http://orcid.org/0000-0002-1564-8120 / Universidade de Coimbra, Portugal

Abstract

Multi-source feedback (MSF) in performance appraisal (PA) enhances fair and objective PA processes. This literature review describes and characterizes the empirical research on MSF in PA processes or as a measurement tool, and proposes new and future research directions considering the gaps found. Using B-ON, EBSCO host and ProQuest and the keywords 'multisource or multi-source', 'performance', and 'appraisal or evaluation or assessment', 162 articles were obtained. After applying the inclusion and exclusion criteria, 37 articles were retained and analyzed. The results demonstrate that the implementation of MSF in PA has increased. It has a political component, allowing organizations to get closer to the concept of decent work and the 2030 Agenda, and a strategic component, enabling employees to increase their performance. In-depth knowledge of MSF in PA requires further studies on its psychometric properties; its cultural sensitivity; links with organizational strategy, policy, and philosophy; and long-term impact.

Keywords: multisource feedback, performance appraisal, literature review.

Investigação Empírica sobre Feedback de Múltiplas Fontes na Avaliação de Desempenho: Revisão de Literatura

Resumo

O feedback de múltiplas fontes (MSF) na avaliação de desempenho (AD) incrementa processos de AD objetivos e justos. Esta revisão de literatura descreve e caracteriza a investigação empírica sobre o MSF nos processos de AD ou como ferramenta de medição e propõe novas e futuras orientações de investigação tendo em conta as lacunas na literatura revista. Recorrendo à B-ON, EBSCO host e ProQuest e às palavras-chave: 'm*ultisource or multi-source*', '*performance*', '*appraisal or evaluation or assessment*', obtiveram-se 162 artigos. Após a aplicação dos critérios de inclusão e exclusão, 37 artigos foram retidos e analisados. Os resultados demonstram que a implementação do MSF na AD aumentou. Este tem uma componente política, permitindo às organizações aproximarem-se do conceito de trabalho digno e da agenda 2030, e uma componente estratégica, possibilitando aos colaboradores incrementar o seu desempenho. O conhecimento aprofundado do MSF na AD carece de mais estudos sobre as suas propriedades psicométricas, a sua sensibilidade cultural, ligação com a estratégia organizacional, política e filosofia, e impacto a longo prazo.

Palavras-chave: feedback de múltiplas fontes, avaliação de desempenho, revisão de literatura.

Investigación Empírica sobre Multisource Feedback en la Evaluación del Desempeño: Revisión de la Literatura

Resumen

La retroalimentación de múltiples fuentes (MSF – *Multisource Feedback*) en la evaluación del desempeño (ED) mejora los procesos de ED objetivos y justos. Esta revisión de la literatura describe y caracteriza la investigación empírica sobre MSF en procesos de ED o como una herramienta de medición y propone nuevas y futuras direcciones de investigación teniendo en cuenta las brechas en la literatura revisada. Utilizando B-ON, EBSCO host y ProQuest y las palabras clave: '*multisource or multi-source*', '*performance*', '*appraisal or evaluation or assessment*', se obtuvieron 162 artículos. Tras aplicar los criterios de inclusión y exclusión, se conservaron y analizaron 37 artículos. Los resultados demuestran que la implementación de MSF en ED ha aumentado. Esto tiene un componente político, que permite a las organizaciones acercarse al concepto de trabajo digno y de la agenda 2030, y un componente estratégico, que permite a los empleados incrementar su desempeño. El conocimiento profundo de MSF en ED carece de más estudios sobre sus propiedades psicométricas, su sensibilidad cultural, su vínculo con la estrategia organizacional, la política y la filosofía, y su impacto a largo plazo.

Palabras clave: retroalimentación de múltiples fuentes, evaluación del desempeño, revisión de la literatura.

Performance appraisal (PA), either informal or formal, is the process that allows organizations to assess employees' performance fairly, for development purposes or to justify decisions, for example. A formal PA with high quality contributes to organizations' development since it provides objectivity, transparency and fairness, interconnecting several human resource practices (dos Santos, 2011) and contributing to a required ethical organizational behavior (Francis, Mónico, Pais, & dos Santos, 2018; Magalhães, dos Santos, & Pais, 2019). A multisource PA system (MSF) refers to formal appraisal of an employee's performance from more than one source, and communication of the results. Its use has increased since the 90s, and sources may be internal (e.g., subordinates, peers, self or supervisors) or external (e.g., clients or patients) (Atwater, Waldman, & Brett, 2002; Loredana & Mirabela, 2015). MSF is used for development purposes and, its positive impact and advantages have been widely and positively recognized by managers (Atwater et al., 2002; Entrekin & Chung, 2001). MSF is seen as more objective, fair, reliable, and valid when compared with traditional PA – someone's performance rated only by his/her supervisor (Baruch & Harel, 1993; Görün et al., 2018). The quality of PA increases when several appraisals are combined (Baruch & Harel, 1993), as supported by empirical research (cf., Weekley & Gier, 1989). Also, several benefits are associated with, but not limited to, total quality management, knowledge of other's perception of one's performance, more transparent, fairer, and informative feedback, and effective communication (Görün, Kayar, & Varol, 2018).

In the current organizational context, described as more ambiguous, dynamic, and demanding, employees are expected to be proactive and autonomous. The present labor context has led to and has emphasized the importance of job crafting, empowering leadership and employee involvement and participation (Rohlfer, 2018; Srinivas & Ashok, 2018). MSF is tuned with these concepts by allowing their correct implementation, facilitation and by enhancing their benefits (Sharma & Kirkman, 2015; Srinivas & Ashok, 2018). Combining viewpoints empowers employees, stimulating their involvement and participation in decisions, feelings of competence, autonomy, and meaning of work (Rohlfer, 2018). By creating conditions for intra-organizational communication and providing information on areas of improvement, MSF nurtures feelings of control and ownership, enhancing the development of the competencies of those involved (Atwater et al., 2002; Sharma & Kirkman, 2015).

MSF is aligned with the decent work (DW) concept. Ferraro, Pais, dos Santos and Moreira (2018) define this through seven dimensions, covering all eleven substantive elements described by the International Labor Organization (ILO). DW is based on human rights at work and on ethical claims (Ferraro, Pais, & dos Santos, 2015), on its insertion on social, work, political and economic context (dos Santos, 2019; Pereira, dos Santos, & Pais, 2019; Ribeiro, 2020), on its ability to express workers' aspirations for their working lives (Ferraro, dos Santos, Pais, & Mónico, 2016) and to promote work engagement (Graça et al., 2019). The fair treatment of employees is one of the components of decent work considering the eleven substantive elements proposed by the ILO and the first dimension of the Ferraro et al. (2018) model. The seventh substantive element is named equal opportunity and treatment in employment (Ferraro et al., 2016). Also the first DW dimension is fundamental principles and values at work (Ferraro et al., 2018). Both are concerned with fair treatment at work and with participation in decisions that impact workers' lives, aligned with the MSF objective: to treat workers fairly (dos Santos, 2011). The 2030 Agenda guides human development until 2030,

consisting of 17 sustainable development goals (SDG). The DW is the eighth objective (United Nations, 2000). A fair PA system is in tune with the requests and considerations of the treaties, declarations, and principles set out above. Therefore, its study, implementation, and practice are crucial to achieving DW and the 2030 goals. PA systems have been studied for several decades (cf. Reinhardt, 1985; Stratton, 1988), but none has focused on MSF. Therefore, besides being more recent, this work will focus on performance appraisal MSF empirical research, aiming to describe and characterize it and propose new and future research directions considering the gaps found.

Method

The PRISMA Protocol is a guide to ensure methodological rigor in systematic reviews in healthcare (Moher et al., 2015). However, we found it inspiring and helpful to adapt to the psychological realm to guide our systematic literature review.

Intending to identify and collect peer-reviewed articles reporting empirical research (no time frame limitation), focused on MSF in Performance Appraisal, published until 31st of December 2019 in academic journals, we searched the following electronic platforms: B-On, EBSCO – host and Discovery, and Proquest. That search included the following academic publication databases: Academic Search Complete, Academic Search Ultimate, American Chemical Society, American Institute of Physics, Annual Reviews, Association for Computing Machinery, Business Source Complete, Business Source Ultimate, Cinahl, Coimbra University Press, Communication Source, Criminal Justice Abstract Full Text, Current Contents (ISI), Dynamed, Ebook Central, eBook Collection (EBSCOhost), EconLit with Full Text, EDS Publication Finder, Education Database, Elsevier, ERIC, Essencial Science Indicators (ISI), Health Business Elite, IEEE, Institute of Physics, ISI Proceedings, Journal Citation Reports (ISI), Library & Information Science Source, LISTA, Medline with full text, Nature, Psychology & Behavioral Science, Psychology Database, Royal Society of Chemistry, Sage, Sage Premier, Society for Industrial and Applied Mathematics, SocINDEX, SocINDEX with Full Text, SPORTDiscuss with Full Text, Springer, Taylor & Francis, Web of Science (ISI), Wiley, and Zentrablatt. These databases were chosen based on the study fields they cover, namely human resources management and related fields.

To retrieve relevant articles the Boolean expressions used were: 'multisource OR multi-source' in abstract AND 'performance' in the title AND 'appraisal OR evaluation OR assessment' in abstract, to avoid literature concerning other forms of assessment or evaluation focus, such as programs, tools or mechanisms. The searches performed retrieved 162 articles.

The articles analyzed were selected after two rounds. In the first round, only one researcher was involved. After assessing each study's relevance through its title and removing duplicates, 85 articles were left to screen. After reading the abstract (and the full-text where it was required) in the second round, some articles were rejected. Specifically, some were not empirical (n=14), were related to another study field (n=30), did not focus on multisource feedback, or did not use it as a measurement tool (n=4). Eventually, 37 articles were kept for more in-depth analysis. In this stage, two researchers were involved plus, a third researcher whenever the first two did not agree on whether or not to include a study. All data was kept and managed on an Excel file.

To meet our study purpose, we sought to identify all studies focused on: a) measuring the relationship between MSF and

one or more variables; b) examining, developing, evaluating, and validating MSF appraisal tools and c) implementing MSF appraisal, and studying its impact on performance appraisal and perceptions towards it. Furthermore, for each study analyzed, we sought to identify the study purpose, data sources, MSF operationalization, the variables measured besides performance, main findings, limitations, suggestions for future research, and practical implications mentioned by the authors (Table 1).

Results

Aiming to be as detailed as possible, we chose to use the word *authors* to refer to the study as a whole and the word *studies* to refer to each study or phase considering that 10 articles include more than one phase (e.g., deLeon & Ewen, 1997).

The study purpose(s) can be grouped in four categories according to its frequency: 1) to examine the relationship between two or more variables, their mediators, moderators or facilitators and the influence, impact or effect of one or more variables, being one of them MSF, into others (20 studies, e.g., Dupee, Ernst, & Caslin, 2011), 2) to examine, develop, evaluate and validate MSF appraisal tools (nine studies, e.g., Bindels et al., 201), 3) to examine MSF psychometric properties (eight studies, e.g., Facteau & Craig, 2001) and 4) to understand how to implement MSF appraisal, its impact on performance and participants' perceptions towards it (four studies, e.g., Lyde, Grieshaber, & Byrns, 2016).

Regarding data sources, sample size ranged from *n* = 9 (Dupee et al. [2011] in post-test sample) to *n* = 2449 (Narayanan, Farmer, & Greco [2018] in Group 1 sample), with the number of participants being mostly in the order of hundreds (31 studies, e.g., Fang, Ko, Chien, & Yu, 2013). Three authors (e.g., Violato, Lockyer, & Fidler, 2008) extracted their samples from existing databases. Twenty-two studies (e.g., Selvarajan & Cloninger, 2012) collected samples from American countries being it's majority from USA (16 studies, e.g., Seaburg et al., 2016), nine (e.g., Jong et al., 2019) from European countries being it's majority from the Netherlands (six studies, e.g., Bindels et al., 2019), nine (e.g., Anand, Hu, Vidyarthi, & Liden, 2018) from Asian countries. Narayanan et al. (2018) used Australian participants. Three authors (e.g., Weigl et al., 2019) did not mention its sample origin.

In 25 studies (e.g., Williams, Byrne, Williams, & Williams, 2017), participants were physicians, in 11 (e.g., Selvarajan & Cloninger, 2012) they were employees in a public or private organization, in six (e.g., Facteau & Craig, 2001) they were managers, in four studies participants were teachers, in four (e.g., Lyde et al., 2016) they were students, in three (e.g., Nuryanti, Nursalam, & Triharini, 2017) they were nurses and in two studies (e.g., Schuh, Zhang, Morgeson, Tian, & van Dick. 2018) participants were engineers.

Performance was rated by internal sources – peers (37 studies, e.g., Bindels et al., 2019), self (31 studies, e.g., Whitaker, 2012), hierarchical superiors (27 studies, e.g., Williams et al., 2017) and subordinates (10 studies, e.g., Hoffman & Woehr, 2009) – and by external sources – patients (six studies, e.g., Overeem et al., 2012a), external observers (two studies, e.g., Gabriel, Acosta, & Grandey, 2015), the organization´s human resource department (three studies, e.g., Treadway et al. [2013] in Study 3) and customers (Gabriel et al., 2015). Performance in Grandey, Houston and Avery's (2018) pilot study was not evaluated. Each author resorted to one (five studies) – e.g., participants declared that were rated by more than one source (e.g., Grandey et al. [2018] in Pilot study) –, a combination of two (14 studies) – e.g., self and peer´s ratings (e.g., Gabriel et al., 2015) –, three (six

studies) – e.g., self, hierarchical superior and peer's ratings (e.g., van Hooft, van der Flier, & Minne, 2006) –, four (16 studies) – e.g., hierarchical superior, peer, subordinate and self-ratings (e.g., Araújo & Taylor, 2012) – or five rating sources (Fang et al., 2013).

MSF was operationalized through questionnaires or forms (33 studies, e.g., Sung & Choi, 2018), tools such as the Child Involvement Scale and Emotional Competence Inventory 2.0 (eight studies, e.g., Jong et al., 2019), portfolios (three studies, e.g., Lyde et al., 2016), single-items (two studies, e.g., Selvarajan & Cloninger, 2012), interviews (two studies, e.g., Overeem et al., 2012a), videotapes or observations (two studies, e.g., Gabriel et al. [2015] in Study 2), self-reflections (Lyde et al., 2016) and human resource department data (e.g., Treadway et al. [2013] in Study 2). Sixteen studies used ad hoc tools (e.g., Noona et al., 2011). Thirty-five studies operationalized MSF through one method, five combined two methods, three combined three methods and one combined four methods. MSF was not operationalized in three studies (e.g., Nuryanti et al. [2017] in Phase 1). Work-related competencies was the most analyzed variable (15 authors, e.g., Gutermann, Lehmann-Willenbrock, Boer, Born, & Voelpel, 2017), followed by demographic variables (14 authors, e.g., Weigl et al., 2019), performance (10 authors, e.g., Gabriel et al., 2015), psychometric properties (10 authors, e.g., Craig & Kaiser, 2003) and individual perceptions and attitudes towards a construct or instruments (five authors, e.g., Anand et al., 2018).

Considering our study purpose, we chose to report the main findings related to MSF. Ten investigations (e.g., Grandey et al., 2018) showed results regarding other variables, which are not performance, its relationships and outcomes. Performance can be predicted, mediated or moderated by the industry type, external demands, training and development, employee competence and commitment and, employees' positive perceptions (e.g., Sung & Choi, 2018). Subjective, objective, contextual, service and in-role performance were positively related with group identity, objective potential diversity, team efficacy and learning, raters' gender and length, perceived warmth, employee gender, conscientiousness, within-group value, organizational citizenship behavior, work-school facilitation, leader-member exchange, emotional and social competences and on-the-job embeddedness (e.g., Araújo & Taylor, 2012).

Regarding MSF appraisal implementation, it was found that although the initial implementation process can be complicated, it becomes easier with successive appraisals (e.g., Noonan, Monagle, & Castanelli, 2011). Also setting size, proper training, clear roles and responsibilities assignment were key to its success being recommended that ratees select their raters (e.g., deLeon & Ewen, 1997).

Regarding MSF psychometric properties, 16 authors (e.g., Violato et al., 2008) revealed that instruments were internally consistent over time. However, there were changes between assessments due to years of practice, gender and practice location.

In general, the self-ratings were the lowest. The rating agreement between supervisor-peer and self-peer was higher than self-subordinate. Moreover, the highest disagreement was found between the self and external source's rating. (e.g., van Hooft et al., 2006). The length of the relationship between ratee and rater, ratee's age, and seniority positively influenced the rating level.

To achieve reliable PA a minimum of two to 15 co-workers, three to nine peers, six to seven hierarchical superiors, and 11 to 25 external sources are recommended (e.g., Overeem et al., 2012b). MSF tools were considered reliable, valid, efficient, usable, portable, comprehensive, highly functional, fair and rich in feedback. Overall, MSF tools were approved by all the involved (e.g., Nuryanti et al., 2017).

Table 1
Study purpose, data source description, multisource operationalization and main findings

Authors and date	Study purpose	Sample dimension (*n*) and description	Multisource operationalization	Main findings of the studies
Anand et al. (2018)	To explore leader-member exchange and within-group value congruence as performance moderators.	*n* = 289 Indian engineers.	Self and managers (questionnaire).	Leader-member exchange and within-group value congruence moderated negatively the link between ideals, in-role performance and organizational citizenship behavior.
Araújo & Taylor (2012)	To determine the influence of emotional and social competence on job performance.	*n* = 36 Peruvian employees.	Supervisors, peers, subordinates and self (EC Inventory 2.0).	Job performance was predicted by emotional and social competencies (ESC). Self-rated ESCs were positively correlated with self-rated transparency.
Bindels et al. (2019)	To evaluate Group Monitor's psychometric properties.	*n* = 254 Dutch physicians.	Staff, peers, self and managers (Group Monitor tool).	Group Monitor internally consistent with good construct validity. Seven to nine staff raters, eight to 15 peers, six to seven managers needed. Four subscales identified for all and a fifth for self and staff.
Buccieri et al. (2008)	Phase 1			
	To examine the assessment tool's content validity.	*n* = 91 American clinical educators.	Not applied.	Raters gave feedback in Administration and Teaching categories. It was preferable to complete the tool annually, electronically or coinciding with each student placement. Center coordinators of clinical education provided more feedback.
	Phase 2			
	To design a tool for directors of clinical educators' use after feedback.	*n* = 591 American directors of physical therapist educators.	Center coordinators of clinical education and instructors (a tool developed).	The tool was comprehensive and could be completed in a reasonable length of time.
Craig & Kaiser (2003)	Phase 1			
	To investigate the consequences of violating statistical assumptions of independent observations.	*n* = 1 000*	Managers in a leadership program.	Violating the independent observations assumption had no effect. A large sample and test had no effects.
	Phase 2			
	Equal to phase 1.	*n* = 200* Equal to phase 1.	Equal to phase 1.	No effect of violating independent observations assumption. A small sample and test had no effects. The chi-square statistics proportion smaller than in phase 1.
Darr & Catano (2008)	To compare MSF ratings with those obtained through a structured behavioral interview.	*n* = 77 Canadian employees.	Supervisors, peers, subordinates and self (ad hoc questionnaire).	Supervisors rated planning and organizing skills higher. Supervisors and peers rated thinking skills higher. Peer-supervisor ratings correlated with the interview. Peer-subordinate correlation was the highest. Ratings obtained through 360° feedback and behavioral interview are convergent.
deLeon and Ewen (1997)	To investigate employees' attitudes towards a multi-source appraisal system.	*n* = 221 (pre-test) *n* = 216 (post-test) American employees.	Peers, self, direct subordinates and supervisor (ad hoc questionnaire).	Employees approved the system and understood it. Raters said the system gave useful information, increased trust, cooperation and motivation. Supervisors valued knowing how workers were perceived.
Dupee et al. (2011)	To investigate MSF's influence on PA satisfaction.	*n* = 11 (pre-test) *n* = 9 (post-test) American nurses.	Supervisor, professional peer, self and workplace peer (questionnaire).	There were no differences between interventions. Workers were satisfied with the appraisal process.
Facteau & Craig (2001)	To test if an MSF instrument shows invariance across rater groups.	*n* = 1 883 American managers.	Self, peer, supervisor and subordinate (multisource appraisal form).	Three items and one scale met the necessary criteria, and all of those were confined to the Motivating Others scale. Latent performance variables varied across all groups.
Fang et al. (2013)	To examine if students enrolled via different programmes had different performances.	*n* = 182 Taiwan medical students.	Self, peer, nursing staff, visiting staff and chief resident (questionnaire).	Teamwork pointed out as the professionalism crucial aspect. Students enrolled through school recommendations were rated higher by them and their peers. Students enrolled through NCUE Examination were rated higher by visiting staff (VS) and chief resident.
Gabriel et al. (2015)	Study 1			
	To understand how service familiarity operates and the impact of employees' positive emotional displays on service performance.	*n* = 114 American employees.	Coworkers and self-report (ad hoc questionnaire).	Employee gender correlated with service performance and occupational service familiarity. Employees' positive emotional displays and service familiarity positively related to coworker ratings. Positive emotional displays positively related to ratings in low-familiarity contexts.
	Study 2			
	Equal to study 1.	*n* = 208 Equal to study 1.	Customers (ad hoc questionnaire + questionnaire), and trained research assistants (videotapes).	Employees' positive emotional displays and service familiarity were predictors of employee friendliness.

Table 1 (continued)
Study purpose, data source description, multisource operationalization and main findings

Authors and date	Study purpose	Sample dimension (*n*) and description	Multisource operationalization	Main findings of the studies
		Pilot Study		
	To verify three assumptions about occupational-racial stereotype incongruence.	*n* = 124 Mostly female and Caucasian.	Not applied.	Occupation predicted occupation warmth. Employee race affected perceived warmth and fit. Men presented more racial bias.
		Study 1		
Grandey et al. (2018)	To verify three assumptions about occupational-racial stereotype incongruence via video stimuli.	*n* = 113 Black and white participants.	Customers (questionnaire).	Employee race affected warmth judgments. Emotional labor improved black participants' evaluations. Respondent race moderated emotional labor.
		Study 2		
	To verify three assumptions about occupational-racial stereotype incongruence.	*n* = 311 Equal do study 1.	Self and supervisors (ad hoc questionnaire).	Interpersonal warmth judgments explained the racial disparity. Expressive regulation improved warmth judgments for black employees.
Gutermann et al. (2017)	To investigate how leaders work engagement spread to followers.	*n* = 88 German team leaders.	Self and subordinates (questionnaire).	Women and older followers reported higher work engagement. Leader-member exchange was positively related to leaders' and followers´ work engagement, which was positively related to performance and negatively to turnover intentions.
Hoffman & Woehr (2009)	1) What are source effect measuring?, and 2) To investigate construct validity for multisource performance rating.	*n* = 404 American managers.	Supervisor, subordinate and peer (performance rating instrument).	Interpersonal skills related to peers. Leadership skills related to the subordinate source. Technical/administrative performance factor related to all sources.
Jong et al. (2019)	To determine if Queen's Simulation Assessment Tool can be used to provide MSF.	*n* = 34 English medicine students.	Self, peer, supervisors and senior experts (ad hoc questionnaire).	Faculty members provided the highest scores. Self-evaluations were the lowest. Physicians and emergency medical service providers agreed the most regarding scores.
Lev & Koslowsky (2012)	To examine if on-the-job embeddedness is a performance predictor.	*n* = 115 Israeli junior/senior-high-school teachers.	Self, school principal and peers (questionnaire).	Conscientiousness predicted task and contextual job performance. On-the-job embeddedness (ONJE) positively related to both performance dimensions. All ONJE components related to contextual performance.
Li & Wong (2008)	To investigate the processes of implementing performance indicators and identify effective practices.	*n* = 284 Japanese teachers.	Self-evaluation portfolios and classroom observations (Child Involvement Scale and Adult Engagement Scale).	Setting size, clear roles and responsibilities were crucial for the implementation of the evaluation and improvement process. Time, parental resistance, space, resources, workload and teachers' motivation were obstacles. Teachers improved in most settings.
Lyde et al. (2016)	To determine faculty's perceptions of the multi-source method.	*n* = 13 American faculty members.	Students and self (student evaluations, instructor portfolio, and a reflection on formative external reviews).	It was beneficial to complete the portfolio, decreasingly so over subsequent submissions. Faculty members were discontent with portfolio timing but approved the tool. The tool was not good enough, and suggestions were made.
		Group 1		
	1) To test the reliability and validity of two MSF questionnaires. 2) To "identify how doctors can use data for professional development purposes".	*n* = 2 449 Australian doctors (MSF as vocational training).	Patients, colleagues and self (patient questionnaire, colleague questionnaire, self-evaluation using colleague and patient questionnaires).	
		Group 2		
Narayanan et al. (2018)	Equal to group 1.	*n* = 1 890 Australian doctors (MSF as development activity).	Equal to group 1.	Communication with patients was positively associated with performance. Self-evaluation was lower. Colleagues-self and patients-self scores converged. Using the patient questionnaire, groups 1 and 4 scored themselves higher. Using the colleague questionnaire, group 4 scored themselves higher. All questionnaires measured different aspects. Patients' and colleagues' questionnaires were valid and reliable.
		Group 3		
	Equal to group 1.	*n* = 375 Australian doctors (compulsory MSF).	Equal to group 1.	
		Group 4		
	Equal to group 1.	*n* = 1 888 Australian doctors (MSF as a regulatory criterion).	Equal to group 1.	

Table 1 (continued)
Study purpose, data source description, multisource operationalization and main findings

Authors and date	Study purpose	Sample dimension (*n*) and description	Multisource operationalization	Main findings of the studies
Noonan et al. (2011)	To develop a multi-source tool.	*n* = 15 Anaesthetists.	Chief anaesthetist, assistants, trainees and chief nurse (ad hoc questionnaire).	The initial process of database development was labour-intensive, but it was easier with successive appraisals. Anaesthetic trainees had the highest non-response rates. In-charge-anaesthetists had the lowest. Assessments were considered valid, reliable and internally consistent.
Nuryanti et al. (2017)	Phase 1			
	To develop a MS performance assessment instrument.	*n* = 53 Indonesian nurses.	Not applied.	The instrument was valid and reliable.
	Phase 2			
	Equal to phase 1.	*n* = 47 Equal to phase 1.	Administrator, head of nurses, self and patients (ad hoc questionnaire).	The instrument was highly functional, reliable, efficient, usable and portable.
Overeem et al. (2012a)	To determine which factors impact on reported change.	*n* = 456 Dutch medical specialists.	Colleagues, coworkers, patients and self (reflective portfolio, interview and questionnaire).	Self-ratings were lower. Performance improved due to MSF assessments. The reported change was positively associated with mentoring quality and web-based service feasibility. Colleagues and self-ratings were negatively correlated with reported change.
Overeem et al. (2012b)	To assess the psychometric properties of three MSF instruments, the influence of sociodemographic variables, the link between self and others' appraisals.	*n* = 146 Dutch physicians.	Peers, non-physician co-workers, patients and self (reflective portfolio, interview and questionnaire).	Peer, co-worker and patient instruments had high internal consistency. Experienced physicians scored lower by peers and co-workers. Male co-workers scored lower. Five peers, five co-workers and 11 patients were required for reliable results. Performance variance was explained by gender and length of the relationship.
Sargeant et al. (2011)	To understand how Nova Scotia Physician Achievement Review was used.	*n* = 23 Canadian family and specialist physicians.	Peers, patients, coworkers and self (ad hoc questionnaire).	To assess Clinical Competence, specialist physicians (SP) used referral letters, and SP and family physicians evaluated the diagnostic tests and pertinence of the medication prescribed.
Schuh et al. (2018)	Study 1			
	To examine employee innovative work behavior and social context effect on performance.	*n* = 29 Chinese engineer team leaders.	Subordinates, colleagues and supervisors (questionnaire)	Innovative work behaviour (IWB) related positively to supervisor ratings and negatively to the leader-member exchange. Supervisors gave a higher rating to employees engaged in IWB.
	Study 2			
	Equal to study 1.	*n* = 177 Chinese full-time employees.	Self (questionnaire).	Employee IWB related positively to supervisor ratings. Leader-member exchange influenced innovative work behaviour and job performance ratings.
Seaburg et al. (2016)	To examine associations between internal medicine residents' publications and their clinical performance.	*n* = 308 American internal medicine residents.	Peers, senior medical residents, faculty and non-physician professionals (questionnaire).	Multi-source evaluations positively linked with the number of scientific publications. Preference to complete the tool annually.
Selvarajan & Cloninger (2012)	To examine the relationship between appraisal characteristics and appraisal outcomes.	*n* = 203 Mexican workers.	Participant´s declaration that had been assessed by more than one source.	MS appraisals perceived as fair. Feedback richness related to procedural and interactive justice. Appraisals used for administrative and developmental purposes perceived as more accurate and fairer. Perception of procedural and interactive justice related to appraisal satisfaction.
Sung & Choi (2018)	To examine the effects of training and development (T&D) on employee outcomes and firms' innovative performance.	*n* = 325* Korean employees, directors and managers.	Self and supervisors (questionnaire).	Firms' innovative performance predicted by the manufacturing industry, market demand, technological change and firm investment in training T&D. Employee competence and commitment predicted firm investment and employee T&D. Positive perceptions of firms' innovative performance when they voluntarily participate in T&D, and firms did not implement it.
Treadway et al. (2013)	Study 1			
	To examine the moderating role of political skill in performance and interpersonal power.	*n* = 97 Canadian employees.	Coworkers (single-item measure) and self-evaluation (18-item Political Skill Inventory).	Age, gender, organizational tenure, performance, political skills and the interaction between them predicted interpersonal power. Employees with higher levels of political skill had more interpersonal power.
	Study 2			
	Equal to study 1.	*n* = 384 American employees.	Organization's human resource department.	Race, position, performance, political skill and interaction between them predicted power. Higher performers influenced the most when their political skill was high.
van der Meulen et al. (2017)	To examine INviting Co-workers to Evaluate Physicians Tool's validity, reliability and feasibility.	*n* = 218 Dutch physicians.	Peers, self, residents and coworkers (ad hoc tool).	Three peers, two to three residents, and three to four coworkers guaranteed reliable assessment. The tool was valid and reliable.

Table 1 (continued)
Study purpose, data source description, multisource operationalization and main findings

Authors and date	Study purpose	Sample dimension (*n*) and description	Multisource operationalization	Main findings of the studies
van Hooft et al. (2006)	To investigate the psychometric properties of MS ratings using Intelligence Test, In-Basket Exercise and Personality Questionnaire.	*n* = 195 Dutch public employees.	Self, supervisor and peer (ad hoc questionnaire).	The agreement was higher for supervisor-peer and self-peer. Supervisor ratings were the lowest or equal to other's ratings. The external construct validity was proven to be weak. Supervisors rated more severely than peers and self.
van Veelen & Ufkes (2019)	To investigate circumstances under which demographic diversity facilitates performance.	*n* = 22 Project teams (Dutch undergraduate PP).	Expert supervisors (ad hoc questionnaire), and self (questionnaire).	Weak group identity scored lower on team task learning, team efficacy and subjective performance. Strong group identity, objective potential diversity and team efficacy positively related to objective and subjective performance.
Violato et al. (2008)	1) To evaluate Physician Achievement Review (PAR) validity and reliability, 2) To investigate changes in performance between assessments, 3) To determine if changes are due to assessments and socio-demographic characteristics.	*n* = 250* Canadian family doctors or general practitioners.	Patients, medical and non-medical colleagues, and self (ad hoc questionnaire).	Instruments were consistent over time. Groups of eight assessors and 25 patients were reliable. There were changes between assessments due to years of practice, gender, place of practice, and communication.
Weigl et al. (2019)	1) To determine the provider's teamwork evaluation, 2) To determine the level of agreement between and within rater and ratee, 3) To determine the link between teamwork and provider outcomes, 4) To determine the link between intra team agreement and provider outcomes.	*n* = 137 Post-surgical providers.	Self and expert-observer (questionnaire and expert observation).	Cooperation and resource management received the lowest ratings. Intra team agreement was better in sub-teams. The self and the observer disagreed in ratings. Stress was higher in sub-teams. Intra team disagreement on stress was related to team performance.
Whitaker (2012)	1) To investigate the task invariance and contextual performance ratings, 2) To examine the effects of covariates on latent task and contextual performance variability.	*n* = 252 American undergraduates, working students.	Supervisor and self (questionnaire).	Supervisor and subordinate's ratings differed. Ratees' age correlated with others' ratings. Ratees' tenure positively associated with self-ratings of task performance and raters' ratings of the task and contextual performance.
Williams et al. (2017)	To determine the level of agreement between a measure (Dreyfus and Dreyfus,1980) and a 360° assessment.	*n* = 264 American physicians.	Leaders, peers, self and reports (questionnaire).	The measure based on the Dreyfus and Dreyfus model (Dreyfus & Dreyfus, 1980) and the 360° instrument had a common factor. There were competencies only captured through 360°.
Wyland et al. (2016)	To examine how work-school interface relates to work outcomes, and how work-school resource demands influence work-school interface.	*n* = 170 American undergraduate, working students.	Self and supervisors (questionnaire).	School-work facilitation positively related to school demands and interpersonal interest/support received at school, and school relevance. School-work conflict positively related to school demands and negatively to task performance. Work-school facilitation positively related to job demands, job control and interpersonal interest/support received at work, and predicted job satisfaction, task performance and interpersonal facilitation. Work-school conflict negatively related to job control and interpersonal interest/support received at work, and positively to job demands. School control negatively related to school-work control.

In general, authors describe MSF developmental benefits, the usefulness of information, and a perceived increase in trust, communication, cooperation, and motivation (e.g., deLeon & Ewen, 1997). In addition, portfolios and mentoring were considered some practices suitable to apply MSF (e.g., Lyde et al., 2016).

Thirty-six authors (e.g., Wyland et al., 2016) highlighted limitations concerning the study design and 16 (e.g., Darr & Catano, 2008) stressed limitations related to the samples used. Thirteen authors (e.g., Craig & Kaiser, 2003) mentioned variables/relationships not included or not explained, and the exclusive focus on certain variables. Thirteen authors (e.g., van Hooft et al., 2006) underlined limitations concerning tools, measures and their psychometric properties. Five authors (e.g., Anand et al., 2018) highlighted limitations related to the data collected. Three authors (e.g., Lyde et al., 2016) did not mention limitations.

Authors proposed further studies on psychometric properties, with more extensive and diverse samples, and expanding the nomological network of MSF (31 authors, e.g., Seaburg et al., 2016). Twenty-two authors (e.g., van Hooft et al., 2006) suggested a future change, re-conceptualization or replication with other conditions of the study procedure, and three (e.g., Buccieri, Brown, Malta, & Ward, 2008) pointed to further development, implementation, and validation of the tools created for that or other contexts. Whitaker (2012) proposed distinguishing between task and contextual performance and what performance means for gender-dominated jobs. Two authors (e.g., Jong et al., 2019) did not suggest future studies.

Regarding practical implications, 25 authors (e.g., Narayanan et al., 2018) emphasized the MSF's role, importance, effects, characteristics, advantages, and the need to include checklists and forms. Its incorporation in organizations was

advised, with regular updates to maintain validity. The use of MSF and a measure based on Dreyfus and Dreyfus's model was recommended (e.g., Williams et al., 2017). Ratees paid more attention to their supervisor's and peers' evaluation, and raters' perceptions were influenced by culture and rating characteristics. Seventeen authors (e.g., van Veelen & Ufkes, 2019) inferred that managers, team leaders and organizations have a crucial role in employees' improvement and in MSF implementation. The inclusion of mentoring in the PA system was suggested. The need to promote programs to develop competencies was pointed out, taking individual variables into account. Two authors (e.g., Facteau & Craig, 2001) highlighted implications regarding the psychometric properties of the MSF tool.

Discussion

MSF in PA has only existed for 30 years (Atwater et al., 2002; Görün et al., 2018). The small number of empirical articles confirms that this subject is still in the early stages of development in psychological research.

MSF's conceptualization, congruent with the democratic ideal, is based on the idea that joining different viewpoints on the same object enriches the whole (Loredana & Mirabela, 2015). The multiplicity of study purposes expresses the growing attempt to understand the various aspects of MSF. Although some improvements have been suggested concerning how MSF should be put into practice, none of the studies concluded that it was useless or had detrimental effects. However, potential adverse effects and disadvantages are associated with using MSF (Hosain, 2016) since some studies found contradictory results in establishing a positive effect (Bailey & Fletcher, 2002). Thus, its potential negative effects should be determined, confirmed, and understood (Table 2).

Knowledge and understanding of MSF and its relationship with other variables are vital to understand the organizational phenomenon, enhance its use and benefits, and intervene suitably in organizations. Future studies should expand the MSF nomological network by researching its relationships with new and diverse variables (Table 2). A small number of empirical investigations assess the psychometric properties of the instruments that allow MSF operationalization, but its usefulness depends on instruments' psychometric properties. Future research should focus on this crucial aspect (Table 2).

MSF responds to organizational, political, and strategic needs. Organizational, due to its recognized advantages and contributions to organizational development (Görün et al., 2018). Political, given its close connection with the DW concept and importance in achieving the 2030 goals (Ferraro et al., 2016). Strategic, as workers' inclusion, empowerment, and involvement in processes such as PA are linked to increased performance (Görün et al., 2018; Rohlfer, 2018).

Additionally, sampling on several continents except Africa reinforces the strong organizational adherence to MSF and the assumption that it is not a context-specific phenomenon. The predominance of North-American samples confirms that they are a major scientific research pole on this subject. The emergence of scientific research in Asia points to the expected progressive importance of research based on that region. There is a lack of studies in African countries (Table 2). Further studies in those countries should be developed since cultural specificities and differences may affect how MSF is perceived and applied and prevent generalization from one cultural context to another (Entrekin & Chung, 2001).

Human resource management practices are most vulnerable to cultural relativity. Thus, practices carried out in one culture may not be easily transferable to another (Entrekin & Chung, 2001). MSF's effectiveness and strength depend on its correct implementation and practice, and attitudes towards it. Further studies are needed to understand MSF's applicability, validity, and effectiveness in eastern culture (Table 2).

MSF, DW, and the 2030 SDG are closely linked since the formers allow the full realization of the latter (Ferraro et al., 2016). MSF in PA is one practice of human resource management. Thus, it contributes extensively to fairer practices, allowing organizations to achieve the DW ideal and the 2030 goals (Ferraro et al., 2016). At the same time, it provides workers

Table 2

Gaps in the existing literature on MSF performance appraisal

Gaps	Suggestions	Assumptions
Lack of empirical studies about MSF.	More empirical studies about MSF.	MSF relevant in the organizational, political and strategic context, being vital to deepen knowledge about it.
Lack of variables related to MSF.	Expanding the MSF nomological network by researching its relationships with new variables.	Knowledge of MSF's relationship with other variables is useful to understand organizational phenomena and intervene in organizations.
Lack of empirical studies focusing on MSF's psychometric properties.	More empirical studies focused on MSF's psychometric properties.	MSF's usefulness depends on its instruments' psychometric properties.
Lack of studies in African countries.	Developing future studies in African countries.	Cultural differences affect the use of MSF.
		Cultural specificities may prevent generalization.
Few studies in non-western cultures.	More studies about the cultural sensitivity of MSF.	Cultural specificities may influence MSF's effectiveness.
		MSF may be more adapted to one culture than to another.
Excessive use of questionnaires in operationalizing MSF.	Operationalization of MSF through different tools.	Data collected through other tools enrich the usefulness of MSF and deepens knowledge about it.
Lack of research on MSF's sensitivity to organizational political factors.	Develop studies investigating MSF's sensitivity to organizational political factors.	Raters may manipulate the appraisals given to respond to their own hidden agenda.
Lack of research about MSF's negative effects at the individual and organizational level.	Develop studies investigating the negative effects of MSF at the individual and organizational level.	MSF might be used as an instrument of moral harassment or raters may have hidden agendas and intentionally manipulate the appraisals.
Lack of studies relating the use of MSF and organizational strategy, policy and philosophy.	Undertaking research relating organizational strategy, policy and philosophy to MSF's inclusion in PA.	Relating various levels of organizational leadership is useful to make organizations more effective and to understand this phenomenon.

and the organization with developmental information, crucial to enhancing performance and organizational development (Görün et al., 2018).

The variety of tools used to operationalize MSF and its adaptation emphasizes the contingency nature of organizations, allowing MSF to be assumed as flexible and adaptable to various organizational contexts (Tureta et al., 2006). These, combined with the advantages found in the literature, justify why its implementation and practice have increased. MSF has been operationalized almost exclusively through questionnaires (Table 2). One of MSF's advantages and objectives is introducing more fairness, objectivity, and validity in PA (Görün et al., 2018). However, operationalizing MSF through instruments other than questionnaires may allow the collection of more diverse data, enriching MSF's usefulness and allowing a deeper understanding of organizational phenomena. Studies that operationalize MSF through different tools are recommended (Table 2).

Employees' strategic involvement and participation in PA are desirable since their empowerment leads to organizational development (Sharma and Kirkman, 2015). Raters have access to partial portions of ratees' performance (Entrekin and Chung, 2001). The combination of viewpoints gives more objectivity and fairness to the process, allowing organizations to get closer to the DW concept and the 2030 Agenda (Atwater et al., 2002; dos Santos, 2019). Self-ratings cannot be considered in isolation when evaluating performance since people tend to underrate themselves (Atwater and Yammarino, 1992). To maintain and/or increase the process's objectivity, validity and fairness, the combination of self-ratings with those of others at different hierarchical levels is recommended.

Organizations are political arenas. There are relationships between organizational, environmental, work, personal factors, and subjective individual agendas (Treadway et al., 2013). Performance management is a complex activity that encompasses many factors, including political considerations, which successfully influence PA. The intentional manipulation of ratings given by supervisors is known to be an organizational phenomenon (Russo, Miraglia, & Borgogni, 2017). MSF brings more objectivity since joining several perspectives reduces the subjectivity inherent to traditional appraisal processes (Görün et al., 2018). MSF can reinforce and support the supervisor's feedback, leading to the predominance of a single viewpoint, thus facilitating the accomplishment of its hidden agendas, goals and wills (Russo et al., 2017; Table 2). Ratees' perceptions of injustice and subjectivity, consequences of subjective and manipulated appraisal, negatively impact at the individual and organizational levels (Görün et al., 2018). Studies about MSF's sensitivity to political factors in organizations and their impact are recommended (Table 2).

The result of PA-MSF is an average of the various ratings given (Görün et al., 2018). MSF offers richer and holistic data about performance compared to with traditional PA, but it dilutes and removes the individual responsibility for intentionally manipulated appraisals (Russo et al., 2017). By not holding raters responsible, MSF can be used as an instrument of moral harassment. This MSF's facet negatively affects employee performance and motivation, leading to stagnation or decline in organizational development. When MSF is used as a means of bullying, it affects its conceptualization as a more reliable, valid, fair, and objective tool since it can be used for personal and evil purposes (Baruch & Harel, 1993). Understanding the extent of this and developing strategies to minimize it is crucial (Table 2).

MSF and its operationalization tools were considered valid, reliable, objective, fair, and consistent over time. Several MSF advantages and benefits were confirmed, cementing the knowledge that already exists. Still, the need for improvements was emphasized. The link between MSF and the current and future political, strategic, organizational, and philosophical context has been confirmed but not investigated (Table 2). Understanding how the various levels of leadership in organizations relate to MSF and how to use it strategically is helpful to make organizations more effective and understand organizational phenomena better. Research relating organizational strategy, policy, and philosophy to MSF is recommended (Table 2).

Limitations and Practical Implications

Our literature review contains some limitations. We only searched for articles written in English, thus excluding publications in other languages. There are other academic publication databases besides those used. Future studies should expand both the languages of publication and databases. Publication bias was not addressed, so other possible studies not published with different results from those presented here were not checked.

MSF in PA is desirable from a strategic, political, and technical viewpoint, its implementation and practice being crucial within the organization. It enhances sustainable organizational development and allows skills enhancement. Plus, managers are vital in this process since they can promote mentoring and appropriate training programs while also maximizing positive attitudes and employee perceptions. MSF is valid and reliable, requiring the use of multiple and diverse sources as much as possible. The authors proposed a minimum number of raters to ensure reliable assessments. Setting size, clear roles, and responsibilities are crucial for correct and successful evaluation and the implementation of improvement processes.

References

Anand, S., Hu, J., Vidyarthi, P., & Liden, R. (2018). Leader-member exchange as a linking pin in the idiosyncratic deals - performance relationship in workgroups. *The Leadership Quarterly, 29*(6), 698–708. https://doi.org/10.1016/j.leaqua.2018.07.005

Araújo, S., & Taylor, S. (2012). The influence of emotional and social competencies on the performance of Peruvian refinery staff. *Cross Cultural Management, 19*(1), 19–29. https://doi.org/10.1108/13527601211195600

Atwater, L., & Yammarino, F. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*, 141-164. https://doi.org/10.1111/j.1744-6570.1992.tb00848.x

Atwater, L., Waldman, D., & Brett, J. (2002). Understanding and Optimizing Multisource Feedback. *Human Resource Management, 41*(2), 193-208. https://doi.org/10.1002/hrm.10031

Bailey, C., & Fletcher, C. (2002). The impact of multiple source feedback on management development: findings from a longitudinal study. *Journal of Organizational Behavior, 23*, 853-867. https://doi.org/10.1002/job.167

Baruch, Y., & Harel, G. (1993), Multi-Source Performance Appraisal: an Empirical and Methodological Note. *Public Administration Quarterly, 17*(1), 96–111. Retrieved from http://www.jstor.org/stable/40862297

Bindels, E., Boerebach, B., van der Meulen, M., Donkers, J., van den Goor, M., Scherpbier, A., Lombarts, K., & Heeneman, S. (2019). A New Multisource Feedback Tool for Evaluating the Performance of Specialty-Specific Physician Groups: Validity of the Group Monitor Instrument. *Journal of Continuing Education in the Health Professions, 39*(3), 168–177. https://doi.org/10.1097/CEH.0000000000000262

Buccieri, K., Brown, R., Malta, S., & Ward, B. (2008). Evaluating the performance of the academic Coordinator/Director of clinical education in physical therapist education: Developing a tool to solicit input from center coordinators of clinical education and clinical instructors. *Journal of Physical Therapy Education, 22*(3), 64-73. Retrieved from https://journals.lww.com/jopte/Abstract/2008/10000/Evaluating_the_Performance_of_the_Academic.8.aspx

Craig, S., & Kaiser, R. (2003). Applying Item Response Theory to Multisource Performance Ratings: What Are the Consequences of Violating the Independent Observations Assumption?. *Organizational Research Methods, 6*(1), 44-60. https://doi.org/10.1177%2F1094428102239425

Darr, W., & Catano, V. M. (2008). Multisource assessments of behavioral competencies and selection interview performance. *International Journal of Selection and Assessment, 16*(1), 68–72. https://doi.org/10.1111/j.1468-2389.2008.00410.x

deLeon, L., & Ewen, A. (1997). Multi-source performance appraisals. *Review of Public Personnel Administration, 17*(1), 22-36. https://doi.org/10.1177/0734371X9701700103

dos Santos, N. R. (2011). O desempenho e a sua avaliação: dicotomias na prática, desafios para a investigação. In A. D. Gomes (Ed.). *Psicologia das organizações, do trabalho e dos recursos humanos* (pp. 577-613). Imprensa da Universidade de Coimbra, Coimbra. Retrieved from https://digitalis.uc.pt/handle/10316.2/31244

dos Santos, N. R. (2019). Decent work expressing universal values and respecting cultural diversity: propositions for intervention. *Psychologica, 62*(1), 233-250. https://doi.org/10.14195/1647-8606_62-1_12

Dupee, J., Ernst, N., & Caslin, E. (2011). Recruitment and Retention Report. Does multisource feedback influence performance appraisal satisfaction?. *Nursing Management, 42*(3), 12–16. https://doi.org/10.1097/01.numa.0000394060.41121.ce

Entrekin, L., & Chung, Y. (2001). Attitudes towards different sources of executive appraisal: a comparison of Hong Kong Chinese and American managers in Hong Kong. *International Journal of Human Resource Management, 12*(6), 965–987. https://doi.org/10.1080/713769693

Facteau, J., & Craig, S. (2001). Are Performance Appraisal Ratings From Different Rating Sources Comparable?. *Journal of Applied Psychology, 86*(2), 215–227. Retrieved from https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1037%2F%2F0021-9010.86.2.215

Fang, J., Ko, Y., Chien, C., & Yu, K. (2013). Assessing medical students' performance in core competencies using multiple admission programs for colleges and universities: From the perspective of multi-source feedback. *Biomedical Journal, 36*(4), 188-197. https://doi.org/10.4103/2319-4170.113856

Ferraro, T., dos Santos, N. R., Pais, L., & Mónico, L. (2016). Historical landmarks of decent work. *European Journal of Applied Business Management, 2*(1), 77-96. Retrieved from https://nidisag.isag.pt/index.php/IJAM/article/view/129

Ferraro, T., Pais, L., & dos Santos, N. R. (2015). Decent work: An aim for all made by all. *International Journal of Social Sciences, IV*(3), 30–42. https://doi.org/10.20472/SS2015.4.3.003

Ferraro, T., Pais, L., dos Santos, N. R., & Moreira, J. (2018). The Decent Work Questionnaire: Development and validation in two samples of knowledge workers. *International Labour Review, 157*(2), 243-265. https://doi.org/10.1111/ilr.12039

Francis, A. G., Mónico, L., Pais, L., & dos Santos, N. R. (2018). Business ethics: a study of Portuguese social representation of business ethics. *International Journal of Business Governance and Ethics, 13*(1), 201, 85-106. https://doi.org/10.1504/IJBGE.2018.095415

Gabriel, A., Acosta, J., and Grandey, A. (2015). The Value of a Smile: Does Emotional Performance Matter More in Familiar or Unfamiliar Exchanges?. *Journal of Business and Psychology, 30*(1), 37–50. https://doi.org/10.1007/s10869-013-9329-2

Görün, M., Kayar, İ., & Varol, B. (2018). 360-Degree Performance Appraisal and Feedback System: A Study with Heads of Departments in Çanakkale Onsekiz Mart University. *Gaziantep University Journal of Social Sciences, 17*(4), 1425–1437. https://doi.org/10.21547/jss.449154

Graça, M., Pais, L., Mónico, L., dos Santos, N. R., Ferraro, T., & Berger, R. (2019). Decent Work and Work Engagement: A Profile Study with Academic Personnel. *Applied Research in Quality of Life.* https://doi.org/10.1007/s11482-019-09780-7

Grandey, A., Houston III, L., & Avery, D. (2018). Fake It to Make It? Emotional Labor Reduces the Racial Disparity in Service Performance Judgments. *Journal of Management, 20*(10), 1-30. https://doi.org/10.1177%2F0149206318757019

Gutermann, D., Lehmann-Willenbrock, W., Boer, D., Born, M., & Voelpel, S. (2017). How Leaders Affect Followers' Work Engagement and Performance: Integrating Leader−Member Exchange and Crossover Theory. *British Journal of Management, 28*(2), 299–314. https://doi.org/10.1111/1467-8551.12214

Hoffman, B., & Woehr, D. (2009). Disentangling the Meaning of Multisource Performance Rating Source and Dimension Factors. *Personnel Psychology, 62*, 735–765. https://doi.org/10.1111/j.1744-6570.2009.01156.x

Hosain, S. (2016). 360 Degree Feedback as a Technique of Performance Appraisal: Does it Really Work?. *Asian Business Review, 6*(13), 21-24. https://doi.org/10.18034/abr.v6i1.779

Jong, M., Elliott, N., Nguyen, M., Goyke, T., Johnson, S., Cook, M., . . . & Kane, B. (2019). Assessment of Emergency Medicine Resident Performance in an Adult Simulation Using a Multisource Feedback Approach. *Western Journal of Emergency Medicine, 2*(1), 64–70. https://doi.org/10.5811/westjem.2018.12.39844

Lev, S., & Koslowsky, M. (2012). On-the-job embeddedness as a mediator between conscientiousness and school teachers' contextual performance". *European Journal of Work and Organizational Psychology, 21*(1), 57–83. https://doi.org/10.1080/1359432X.2010.535656

Li, H., & Wong, N. (2008). Implementing Performance Indicators of early learning and teaching: A Chinese study. *International Journal of Early Years Education, 16*(2), 115–131. https://doi.org/10.1080/09669760701516942

London, M., & Smither, J. (1995). Can Multi-Source Feedback Change Perceptions of Goal Accomplishment, Self-Evaluations, and Performance-Related Outcomes?. Theory-Based Applications and Directions for Research. *Personnel Psychology, 48*(4), 803–839. https://doi.org/10.1111/j.1744-6570.1995.tb01782.x

Loredana, L., & Mirabela, F. (2015). Multi-Rater Feedback - in Promoting and Stimulating an Institution or Organization Employees. *Annals of the "Constantin Brâncuşi" University of Târgu Jiu, Economy Series*, 399–403. Retrieved from https://econpapers.repec.org/article/cbujrnlec/y_3a2015_3av_3aspecial_3ap_3a399-403.htm

Lyde, A., Grieshaber, D., & Byrns, G. (2016). Faculty Teaching Performance: Perceptions of a Multi-Source Method for Evaluation (MME). *Journal of the Scholarship of Teaching and Learning, 16*(3), 82–94. Retrieved from https://ir.library.illinoisstate.edu/sotlpubs/92

Magalhães, A., dos Santos, N. R., & Pais, L. (2019). Multi-source research designs on ethical leadership: A literature review. *Business and Society Review, 124*(3), 345-364. https://doi.org/10.1111/basr.12179

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . & Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*(1). https://doi.org/10.1186/2046-4053-4-1

Narayanan, A., Farmer, E., & Greco, M. (2018). Multisource feedback as part of the medical board of Australia's professional performance framework: Outcomes from a preliminary study. *BMC Medical Education, 18*, 323-334. https://doi.org/10.1186/s12909-018-1432-7

Noonan, C., Monagle, J., & Castanelli, D. (2011). Development of a multi-source feedback tool for consultant anaesthetist performance. *Australian Health Review, 35*(2),141–145. https://doi.org/10.1071/AH10928

Nuryanti, A., Nursalam, N., & Triharini, M. (2017). Development of Performance Assessment Instrument for Nurses Based on Web in Inpatient Unit. *Jurnal Ners, 12*(1), 108-112. https://doi.org/10.20473/jn.v12i1.2253

Overeem, K., Wollersheim, H., Arah, O., Cruijsberg, J., Grol, R., & Lombarts, K. (2012b). Evaluation of physicians' professional performance: An iterative development and validation study of multisource feedback instruments. *BMC Health Services Research, 12*(80). https://doi.org/10.1186/1472-6963-12-80

Overeem, K., Wollersheimh, H., Arah, O., Cruijsberg, J., Grol, R., & Lombarts, K. (2012a). Factors predicting doctors' reporting of performance change in response to multisource feedback. *BMC Medical Education, 12*(52). https://doi.org/10.1186/1472-6920-12-52

Pereira, S., dos Santos, N. R., & Pais, L. (2019). Decent work's contribution to the economy for the common good. *International Journal of Organizational Analysis*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJOA-07-2019-1840

Reinhardt, C. (1985). The State of Performance Appraisal: A Literature Review. *Human Resource Planning, 8*(2), 105-110.

Ribeiro, M. A. (2020). Contribuições da psicologia para repensar o conceito de trabalho decente. *Revista Psicologia Organizações e Trabalho, 20*(3), 1114-1121. https://doi.org/10.17652/rpot/2020.3.19488

Rohlfer, S. (2018). Employee involvement and participation in SMEs: a synthesis of extant research. *Journal of Evolutionary Studies in Business, 3*(1), 112-136. Retrieved from https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1344%2Fjesb2018.1.j039

Russo, S., Miraglia, M., & Borgogni, L. (2017). Reducing Organizational Politics in Performance Appraisal: The Role of Coaching Leaders for Age-diverse Employers. *Human Resource Management, 56*(5), 769-783. https://doi.org/10.1002/hrm.21799

Sargeant, J., Macleod, T., Sinclair, D., & Power, M. (2011). How do physicians assess their family physician colleagues' performance? Creating a rubric to inform assessment and feedback. *Journal of Continuing Education in the Health Professions, 31*(2), 87–94. https://doi.org/10.1002/chp.20111

Schuh, S., Zhang, X., Morgeson, F., Tian, P., & van Dick, R. (2018). Are you really doing good things in your boss's eyes? Interactive effects of employee innovative work behavior and leader–member exchange on supervisory performance ratings. *Human Resource Management, 57*(1), 397–409. https://doi.org/10.1002/hrm.21851

Seaburg, L., Wang, A., West, C., Reed, D., Halvorsen, A. J., Engstler, G., Oxenteko, A., & Beckman, T. (2016). Associations between resident physicians' publications and clinical performance during residency training. *BMC Medical Education, 16*(22). https://doi.org/10.1186/s12909-016-0543-2

Selvarajan, T., & Cloninger, P. (2012). Can performance appraisals motivate employees to improve performance? A Mexican study. *International Journal of Human Resource Management, 23*(15), 3063-3084. https://doi.org/10.1080/09585192.2011.637069

Sharma, P., & Kirkman, B. (2015). Leveraging Leaders: A Literature Review and Future Lines of Inquiry for Empowering Leadership Research. *Group and Organization Management, 40*(2), 193–237. https://doi.org/10.1177/1059601115574906

Srinivas, S., & Ashok, K. (2018). The Role of Job Crafting in Feedback-Performance Relationship: A Study of Supervisors and Managers of the Indian Railways. *IUP Journal of Organizational Behavior, 17*(4), 58–76. Retrieved from https://ssrn.com/abstract=3381443

Stratton, K. (1988). Performance appraisal and the need for an organizational grievance procedure: A review of the literature and recommendations for future research. *Employee Responsibilities and Rights Journal, 1*(3), 167–179. https://doi.org/10.1007/BF01384979

Sung, S., & Choi, J. (2018). Effects of training and development on employee outcomes and firm innovative performance: Moderating roles of voluntary participation and evaluation. *Human Resource Management, 57*(6), 1339–1353. https://doi.org/10.1002/hrm.21909

Treadway, D., Breland, J., Williams, L., Cho, J., Yang, J., & Ferris, G. (2013). Social Influence and Interpersonal Power in Organizations: Roles of Performance and Political Skill in Two Studies. *Journal of Management, 39*(6), 1529–1553. https://doi.org/10.1177/0149206311410887

Tureta, C., Rosa, A., & Ávila, S. (2006). Da Teoria Sistêmica ao Conceito de Redes Interorganizacionais: Um Estudo Exploratório da Teoria das Organizações. *Revista de Administração da UNIMEP, 4*(1), 1-16. https://doi.org/10.15600/1679-5350/rau.v4n1p1-16

United Nations. (2000). *United Nations Millennium Declaration: Resolution 55/2 adopted by the United Nations General Assembly on 8 Sep. 2000 Declaration.* Retrieved from http://www.un.org/mil-lennium/declaration/ares552e.pdf

van der Meulen, M., Boerebach, B., Smirnova, A., Heeneman, S., oude Egbrink, M. G. A., Van der Vleuten, C., Arah, O., & Lombarts, K. (2017). Validation of the INCEPT: A Multisource Feedback Tool for Capturing Different Perspectives on Physicians' Professional Performance. *Journal of Continuing Education in the Health Professions, 37*(1), 9–18. https://doi.org/10.1097/ceh.0000000000000143

van Hooft, E., van der Flier, H., and Minne, M. (2006). Construct Validity of Multi-Source Performance Ratings: An Examination of the Relationship of Self-, Supervisor-, and Peer-Ratings with Cognitive and Personality Measures. *International Journal of Selection and Assessment, 14*(1), 67–81. https://doi.org/10.1111/j.1468-2389.2006.00334.x

van Veelen, R., & Ufkes, E. (2019). Teaming Up or Down? A Multisource Study on the Role of Team Identification and Learning in the Team Diversity–Performance Link. *Group and Organization Management, 44*, 38–71. https://doi.org/10.1177/1059601117750532

Violato, C., Lockyer, J., & Fidler, H. (2008). Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme. *Medical Education, 42*, 1007–1013. https://doi.org/10.1111/j.1365-2923.2008.03127.x

Weekley, J., & Gier, J. (1989). Ceiling in the Reliability and Validity of Performance Ratings: The Case of Expert Raters. *Academy of Management Journal, 32*, 213-222. https://doi.org/10.5465/256428

Weigl, M., Heinrich, M., Keil, J., Wermelt, J. Z., Bergmann, F., Hubertus, J., & Hoffmann, F. (2019). Team performance during postsurgical patient handovers in paediatric care. *European Journal of Pediatrics*, 1-10. https://doi.org/10.1007/s00431-019-03547-w

Whitaker, B. (2012). The Influence of Covariates on Latent Job Performance Ratings from Differing Sources: A Mimic Modeling Approach. *International Journal of Organization Theory and Behavior, 15*(2), 176–198. https://doi.org/10.1108/IJOTB-15-02-2012-B002

Williams, B., Byrne, P, Williams, N., & Williams, M. (2017). Dreyfus and Dreyfus and Indicators of Behavioral Performance: A Study of Measurement Convergence. *Journal of Continuing Education in the Health Professions, 37*(1), 50–54. https://doi.org/10.1097/ceh.0000000000000138

Wyland, R., Lester, S., Ehrhardt, K., & Standifer, R. (2016). An Examination of the Relationship Between the Work-School Interface, Job Satisfaction, and Job Performance. *Journal of Business and Psychology, 31*(2), 187–203. https://doi.org/10.1007/s10869-015-9415-8

Information about the authors:

Mariana Lobo Moreno
Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra
Rua do Colégio Novo, s/n
3001-802 Coimbra, Portugal
E-mail: marianamoreno97@hotmail.com

Nuno Rebelo dos Santos
E-mail: nrs@uevora.pt

Leonor Maria Pais
E-mail: leonorpais@fpce.uc.pt