



Differences in self- and managerial-ratings on generic performance dimensions

Authors:

Xander van Lill^{1,2}
Gerda van der Merwe³

Affiliations:

¹Department of Industrial Psychology and People Management, College of Business and Economics, University of Johannesburg, Johannesburg, South Africa

²Department of Product and Research, JVR Africa Group, Johannesburg, South Africa

³JVR Consulting, JVR Africa Group, Johannesburg, South Africa

Corresponding author:

Xander van Lill,
xvanlill@gmail.com

Dates:

Received: 22 July 2022
Accepted: 12 Sept. 2022
Published: 29 Nov. 2022

How to cite this article:

Van Lill, X., & Van der Merwe, G. (2022). Differences in self- and managerial-ratings on generic performance dimensions. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 48(0), a2045. <https://doi.org/10.4102/sajip.v48i0.2045>

Copyright:

© 2022. The Authors.
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Orientation: The 360-degree performance assessments are frequently deployed. However, scores by different performance reviewers might erroneously be aggregated, without a clear understanding of the biases that are inherent to different rating sources.

Research purpose: The purpose of this study was to determine whether there are conceptual and mean score differences between self- and managerial-ratings on performance dimensions.

Motivation for the study: Combining self- and managerial-ratings may lead to incorrect decisions about the development, promotion, and/or remuneration of employees. Understanding the effects of rating sources may aid thoughtful decisions about the applications of self- versus managerial-ratings in low- and high-stakes decisions.

Research approach/design and method: A cross-sectional design was implemented by asking 448 managers to evaluate their subordinates' performance, and 435 employees to evaluate their own performance. The quantitative data were analysed by means of multi-group factor analyses and robust *t*-tests.

Main findings: There was a satisfactory degree of structural equivalence between self- and managerial-ratings. Practically meaningful differences emerged when the means of self- and managerial-ratings were compared.

Practical/managerial implications: It might be meaningful to uncouple self- and managerial-ratings, when providing performance feedback. Managerial ratings might be a more conservative estimate, which could be used for high-stakes decisions, such as remuneration or promotion.

Contribution/value-add: This study is the first to investigate the effect of rating sources on a generic model of performance in South Africa. It provides valuable evidence regarding when different rating sources should be used in predictive studies, performance feedback, or high-stakes talent decisions.

Keywords: individual work performance; generic performance; performance measurement; rating sources; 360-degree performance feedback.

Introduction

Orientation

A function central to management is motivating employees and directing their behaviours towards achieving collective goals (Aguinis, 2019). Within-group cohesion based on collective goals is, in turn, paramount to maintain organisational effectiveness in a competitive business landscape (Campbell & Wiernik, 2015; Hogan & Sherman, 2020). A key feature of the managerial function is the continuous review of, and feedback on, performance, to reinforce enterprising behaviour and to calibrate actions against larger goals. Performance management is a perplexing area of practice, and seems to give rise to several management fads, often advertised as simple solutions to the complex problem of increasing desirable work behaviours (Pulakos et al., 2019). Such simple answers tend to deviate from evidence-based practices, and might do nothing to increase performance, or even decrease performance (Pulakos et al., 2019; Rousseau, 2012).

The complexity of performance management could be attributable to its multidisciplinary nature and the diverse areas of research it taps into, which include theories of measurement, individual differences, work motivation, as well as cognitive and behavioural psychology (Aguinis, 2019; Pulakos et al., 2019). A practice that took hold in the 1980s, and which has

gained popularity in performance reviews since, is multisource ratings or 360-degree ratings performed by, for example, managers, peers, subordinates, clients, and employees (Aguinis, 2019; Pulakos et al., 2019). It is theorised that pooling ratings from different sources is likely to provide a more comprehensive and reliable review of an employee's performance (Harari & Viswesvaran, 2018). It is also argued that the degree of self-other agreement or disagreement in performance ratings may be especially valuable when providing multisource feedback for development purposes (Heidemeier & Moser, 2009).

Hogan and Sherman (2020) argued that it is a social skill to translate identity (employees' ratings of their own performance, which is subjective) into reputation (objective ratings of employees' performance by others). This is argued to enable differentiation of dysfunctional employees from competent employees. Employees who are dysfunctional, possibly because of low self-esteem (Atwater & Yammarino, 1992), might have distorted views of their own performance, and may be less willing to change their behaviours based on others' feedback. Competent employees, by contrast, effectively integrate performance feedback and adjust their current self-perceptions – and their resultant behaviour – to improve their reputations at work (Hogan & Sherman, 2020).

Trust in multisource feedback, especially when there are disagreements between self ratings and other ratings, might be fickle if there is no clear evidence that the dimensions being assessed are interpreted the same across different performance reviewers (Heidemeier & Moser, 2009; Scullen et al., 2003). Furthermore, cognitive biases, such as leniency bias, might lead to significant mean group differences between self ratings and other ratings. Employees are perhaps more likely to give themselves higher performance ratings than, for example, what their managers would grant (Aguinis, 2019). Establishing equivalence in rating perspectives and inspecting mean group differences between rating sources are, therefore, crucial steps in validating and determining the relevance of separate norms for work performance measures.

Performance management in South Africa has a rich scientific history in terms of the reliable and valid measurement of generic general, broad, and narrow dimensions of individual work performance (Van Lill & Taylor, 2022). However, no published scientific research has been conducted on the equivalence and mean group differences of self ratings versus other ratings based on generic performance measures developed locally. One reason is the reliance on single-source performance reviews in South Africa, mainly either self- or managerial-ratings (Myburgh, 2013; Schepers, 2008; Van der Vaart, 2021; Van Lill & Taylor, 2022). Evidence-based practises around performance development in South Africa, especially performance feedback on the agreement or disagreement between self ratings versus other ratings, require further empirical scrutiny.

Purpose of the study

Van Lill and Taylor (2022) highlighted the effect of rater sources as one of the shortcomings of current research on the Individual Work Performance Review (IWPR). The purpose of this study was to inspect the measurement invariance and mean group differences between self- and managerial-ratings. In doing so, it could be established whether employees and managers have the same conceptual understanding of performance when conducting reviews, and help people practitioners and managers to understand possible biases that might be at play within a particular rating source.

Literature review

The theoretical frame underlying the IWPR was utilised to structure the present study (Van Lill & Taylor, 2022). The theoretical framework outlines five broad performance dimensions, identified by Van Lill and Taylor (2022): *in-role*, *extra-role*, *adaptive*, *leadership*, and *counterproductive performance*. According to Van Lill and Taylor (2022, p. 3–5):

1. *In-role performance* refers to: 'Actions that are official or known requirements for employees (Carpini et al., 2017; Motowidlo & Van Scotter, 1994). These behaviours could be viewed as the technical core (Borman & Motowidlo, 1997) that employees must demonstrate to be perceived as proficient and able to contribute to the achievement of organisational goals' (Carpini et al., 2017).
2. *Extra-role performance* refers to: 'future- or change-orientated acts (Carpini et al., 2017), aimed at benefitting co-workers and the team (Organ, 1997), that are discretionary or not part of the employee's existing work responsibilities' (Borman & Motowidlo, 1997).
3. *Adaptive performance* relates to: 'employees' demonstration of the ability to cope with and effectively respond to crises or uncertainty' (Carpini et al., 2017; Pulakos et al., 2000).
4. *Leadership performance* refers to: 'the effectiveness with which an employee can influence co-workers to achieve collective goals' (Campbell & Wiernik, 2015; Hogan & Sherman, 2020; Yukl, 2012).
5. *Counterproductive performance* reflects on the: 'intentional or unintentional acts (Spector & Fox, 2005) by an employee that negatively affect the effectiveness with which an organisation achieves its goals and cause harm to its stakeholders' (Campbell & Wiernik, 2015; Marcus et al., 2016).

These five broad performance dimensions are further theorised to break down into 20 narrow performance dimensions, as shown in Figure 1. Definitions of the narrow dimensions are outlined in Table 1. As portrayed in Figure 1, it is theorised that a general factor stands at the apex of all the performance dimensions identified in the IWPR. A study conducted by Van Lill and Van der Vaart (2022) replicated Viswesvaran et al.'s (2005) discovery of a general performance factor in South Africa. To this end, Van Lill and Van der Vaart (2022) forwarded preliminary evidence,

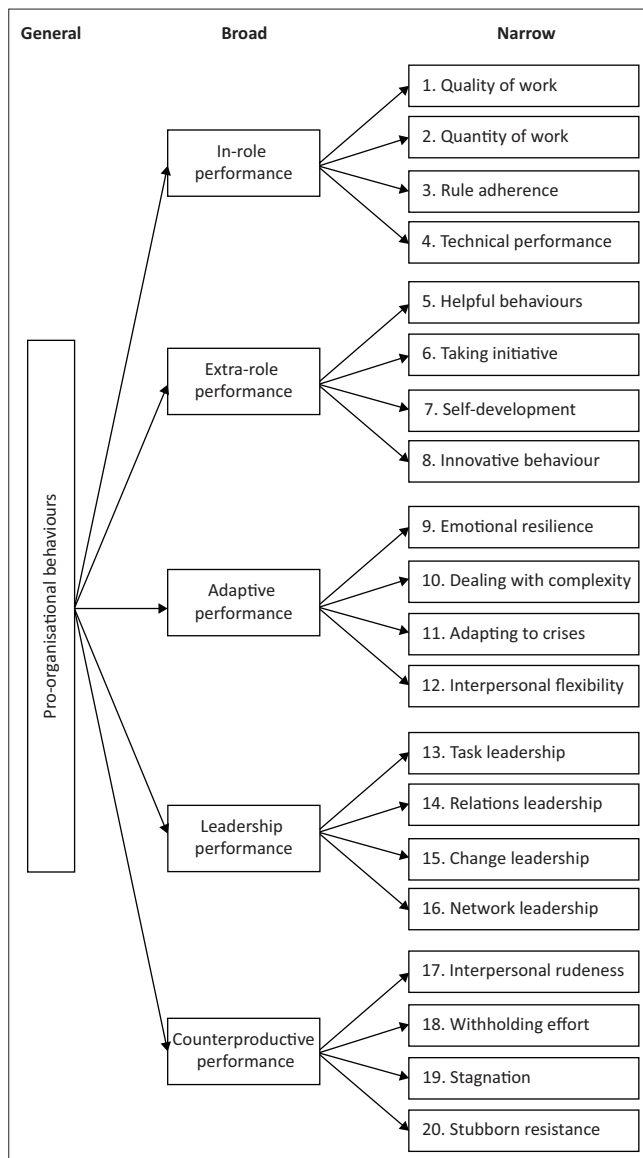


FIGURE 1: Performance levels of the individual work performance review (IWPR).

suggesting that a large amount of common variance among the 20 narrow performance dimensions in the IWPR is explained by a general factor of performance. The present study focused only on the measurement invariance and mean group differences among the broad and narrow dimensions, but it was theorised that the strong general factor might explain common trends among all the performance dimensions.

Measurement equivalence of performance dimensions

A first important step in the present study, before mean group differences between different rating sources were inspected, was to determine whether employees and managers have the same conceptual understanding of the performance dimensions measured. One method to determine whether different groups of raters have the same understanding of the theoretical structure underlying the performance questions asked is the test of measurement invariance (Scullen et al., 2003; Vandenberg & Lance, 2000). A meta-analytical study

TABLE 1: Definitions of the narrow performance dimensions in the individual work performance review (IWPR).

| Dimension | Definition |
|---------------------------------|--|
| Quality of work | The thoroughness with which employees perform work tasks, evident in the degree to which employees pay attention to detail and minimise errors |
| Quantity of work | How productive employees are in meeting challenging work goals in terms of both the volume of output and meeting the required time frame |
| Rule adherence | Employees' tendency to comply with informal and formal rules and regulations of the organisation |
| Technical performance | The degree to which employees perform well at tasks that are differentiated, complicated, and require a certain level of expertise |
| Helpful behaviours | Employees' acts of kindness towards co-workers |
| Taking initiative | Demonstrated by employees showing self-starting behaviour and doing more than is expected of them |
| Self-development | Reflected in employees' initiatives to enhance their competence by actively gaining knowledge and learning new skills that could benefit the team |
| Innovative behaviours | Employees exploring or generating new opportunities and implementing new and creative ideas |
| Emotional resilience | Demonstrated when employees maintain their composure when they have to work under high pressure |
| Dealing with complexity | Demonstrated when employees think, decide, and act sensibly under uncertain and unusual situations when there are no clear guidelines |
| Adapting to crises | The degree to which employees remain objective, make swift decisions, and react with appropriate urgency to a crisis |
| Interpersonal flexibility | Reflected in how comfortable employees are with situations in which people with diverse views do not agree with each other; it is also represented by employees' open-mindedness in interaction with co-workers from different backgrounds |
| Task-orientated leadership | Demonstrated by employees when they direct the efforts of co-workers towards the achievement of team goals |
| Relations-orientated leadership | Demonstrated when consideration is used to empower and motivate co-workers to achieve team goals |
| Change-orientated leadership | The degree to which employees inspire their co-workers to effect required changes to the way they do their work |
| Network-orientated leadership | The degree to which networking is used to connect co-workers with key role players inside and outside the organisation |
| Interpersonal rudeness | Disrespectful acts that reflect a lack of regard for others |
| Withholding effort | Demonstrated when employees show a lack of enthusiasm in their work by exerting less effort than is expected for the position they hold |
| Stagnation | Demonstrated when an employee displays an unwillingness to learn new skills, thereby affecting team effectiveness |
| Stubborn resistance | Reflected in an employee's unreasonable opposition to change or an unwillingness to support initiatives at work, and suggests a destructive form of opposition to team goals |

Source: Obtained with permission from Van Lill, X., & Taylor, N. (2022). The validity of five broad generic dimensions of performance in South Africa. *South African Journal of Human Resource Management*, 20, 1–15. <https://doi.org/10.4102/sajhrm.v20i0.1844>

conducted by Scullen et al. (2003) revealed, in accordance with the findings of Fecteau and Craig (2001), a sufficient degree of measurement invariance between rating sources to suggest that employees and managers might hold the same conceptual understanding of the generic performance dimensions measured. However, Scullen et al. (2003) indicated that the factor variances of self-ratings were consistently lower compared to managerial ratings. Scullen et al. (2003) also argued that employees might use a smaller range of the construct continuum when rating themselves, compared with managerial ratings. However, when Scullen et al. (2003) relaxed the variances of self-ratings, it appeared that the invariance of the factor models still held up to scrutiny.

The five broad dimensions identified by Van Lill and Taylor (2022) are theorised to be hierarchical (bifactor) models, with

a broad general dimension and four narrower associated dimensions (see Figure 1). It was subsequently hypothesised that:

H₁: The bifactor structure of each of the five broad performance dimensions displays measurement invariance across self-ratings and managerial ratings.

Mean difference in self- and managerial-ratings on performance

A meta-analytical study conducted by Heidemeier and Moser (2009) revealed that self-ratings are consistently lower than managerial ratings. Conway and Huffcutt (1997) further reported low meta-analytically derived inter-rater reliabilities between self- and managerial-ratings. This may be attributable to a leniency bias, where employees attempt to manage the social impressions of others by giving themselves higher ratings. Leniency bias in self-ratings could also be unconsciously performed, via self-deception, as a self-protective mechanism in defence of the ego-threatening nature of performance evaluations. However, Heidemeier and Moser (2009) indicated that *in-role performance* ratings display lower mean group differences between self- and managerial-ratings, compared to other theoretical dimensions of performance. Expectations around *in-role performance*, relative to *extra-role performance*, might be more explicit and, therefore, result in a greater agreement between self- and managerial-ratings.

Based on the meta-analytical evidence of Heidemeier and Moser (2009), it was hypothesised that:

H_{2A}: Employees' self-ratings are more lenient, compared to managerial-ratings, on all five of the broad performance dimensions.

H_{2B}: Employees' self-ratings are more lenient, compared to managerial-ratings, on all 20 of the narrow performance dimensions.

Research design

Research approach

A cross-sectional, quantitative research design was utilised in the current study. A cross-sectional design enabled a nuanced view of the nature of performance at a single point in time and an efficient quantitative exploration of relationships between a large set of variables across different organisational contexts (Spector, 2019). Multiple sources, namely, self-ratings and managerial-ratings, further assisted to expand the cross-sectional design by accounting for a source of method variance (Podsakoff et al., 2012), namely, the leniency bias associated with self-ratings or the halo bias associated with managerial-ratings (Aguinis, 2019; Spector, 2019).

Research method

Research participants

The researchers aimed to increase the generalisability of the study by inviting 15 organisations across different economic sectors to participate (Aguinis & Edwards, 2014; Scullen

et al., 2003). A total of 883 performance ratings of South African employees were completed by a cohort of managers ($n = 448$; 51%) and employees ($n = 435$; 49%) in six participating organisations, via a census (stratified sampling strategy). The final sample represented the following sectors: oil and gas, agriculture, finance, professional services, and information technology. The mean age of employees was 39.05 years (standard deviation [SD] = 7.47 years). Most of the employees self-identified as white ($n = 365$; 46%), followed by black African ($n = 218$; 27%), Indian (158; 20%), Coloured (individuals of mixed ancestry; $n = 48$; 6%), and Asian (5; 1%). The sample comprised more women ($n = 429$; 54%) than men ($n = 366$; 46%). Most of the employees were registered professionals ($n = 270$; 34%), followed by mid-level managers ($n = 200$; 25%), skilled employees (168; 21%), low-level managers ($n = 142$; 18%), semi-skilled employees ($n = 9$; 2%), and top-level managers (6; 1%).

Measuring instrument

The IWPR (Van Lill & Taylor, 2022) was administrated to collect performance data. The IWPR consists of 80 items (four items for each of the 20 narrow performance dimensions), covering five factors. Each item was measured using a five-point behaviour frequency scale (Aguinis, 2019). Word anchors defined the extreme points of each scale, namely, (1) *Never demonstrated* and (5) *Always demonstrated*. Casper et al.'s (2020) guidelines were used to qualitatively anchor numeric values between the extreme points, to better approximate equal empirical intervals on the rating scale, namely, (2) *Rather infrequently demonstrated*, (3) *Demonstrated some of the time*, and (4) *Quite often demonstrated*. The narrow dimensions of the IWPR displayed good internal consistency reliability for self-ratings (α and $\omega \geq 0.68$) and managerial ratings (α and $\omega \geq 0.83$).

Research procedure

Employees and managers completed the review via an e-mail link. All participants received information on the development purpose of the study. The researchers deliberately chose a development focus, as ratings aimed at development tend to be less lenient and more accurate than those used for administrative (promotion or remuneration) purposes (Scullen et al., 2003). Managers and employees were also informed about the nature of the measurement, voluntary participation, benefits of participation, anonymity of their personal data, and that the data would be used for research purposes.

Statistical analysis

Vandenberg and Lance's (2000) recommended sequence for conducting a test of measurement invariance across multiple groups was used in the present study. The equivalence of the bifactor model for each of the five broad performance dimensions was determined. Bifactor models were the best-fitting models identified in the study conducted by Van Lill and Taylor (2022). Measurement invariance was conducted by comparing the fit statistics of a multigroup confirmatory factor model for configural invariance (i.e. equality of the

factor structure across groups), metric invariance (i.e. similar factor loadings across groups), scalar invariance (similar intercepts across groups), and strict invariance (similar residual variances), using Version 0.6-11 of the lavaan package (Rosseel, 2012; Rosseel et al., 2022) in R (R Core Team, 2016).

Mardia's multivariate skewness (190948.90; $p < 0.001$) and kurtosis (220.90; $p < 0.001$) coefficients indicated that the data had a non-normal multivariate distribution. Given the medium ($n = 448$) sample size (Rhemtulla et al., 2012), the employment of rating scales with five numerical categories (Rhemtulla et al., 2012), and violation of multivariate normality (Satorra & Bentler, 1994; Yuan & Bentler, 1998), robust maximum likelihood (MLM) estimation was deemed appropriate (Bandalos, 2014). Model-data fit of the Confirmatory Factor Analysis (CFA) was evaluated using the comparative fit index (CFI), the Tucker-Lewis Index (TLI), standardised root mean-square residual (SRMR), and root mean square error of approximation (RMSEA) (Brown, 2006; Hu & Bentler, 1999). The fit was considered suitable if the RMSEA and SRMR were ≤ 0.08 (Brown, 2006; Browne & Cudeck, 1992) and the CFI and TLI were ≥ 0.90 (Brown, 2006; Hu & Bentler, 1999). Brown (2006) argues that, even if CFIs display a marginally good fit to the data (CFI and TLI in the range of 0.90 to 0.95), models may still be considered to display acceptable fit if other indices (SRMR and RMSEA), in tandem, are within the acceptable range.

After inspecting bias with the test of invariance, robust t -tests were conducted on the IWPR's broad and narrow dimensions with Version 1.1-3 of the WRS2 package in R (Mair & Wilcox, 2020, 2021), to inspect the mean group differences between self- and managerial ratings.

Ethical considerations

Approval to conduct the study was obtained from the Department of Industrial Psychology and People Management (IPPM) Research Ethics Committee, University of Johannesburg (reference no: IPPM-2020-455).

Results

Table 2 reports the inter-factor correlations of the self- and managerial-ratings. The inter-factor correlations were obtained by conducting an oblique lower-order confirmatory factor model. The fit statistics for the oblique lower-order confirmatory factor model of the entire IWPR for self-ratings ($\chi^2 [df] = 4200.34 [2890]$; CFI = 0.92; TLI = 0.91; SRMR = 0.06; RMSEA = 0.04 [0.03; 0.04]) and managerial-ratings ($\chi^2 [df] = 4769.72 [2890]$; CFI = 0.94; TLI = 0.93; SRMR = 0.05; RMSEA = 0.04 [0.04; 0.05]) were satisfactory (Brown, 2015; Hu & Bentler, 1999).

Previous evidence suggests that the narrow dimensions of the IWPR display a sufficient degree of discriminant validity (Van Lill & Taylor, 2022; Van Lill & Van Der Vaart, 2022).

TABLE 2: Latent factor correlations from the 20-factor Confirmatory Factor Analysis for manager ratings (under the diagonal) and self-ratings (above the diagonal).

| DIM | QLW | QNW | REA | TNP | HPB | TII | SFD | IOB | ETR | DLC | APC | IEF | TKL | RAL | CNL | NWL | IER | WHE | SGN | SBR |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| QLW | - | 0.79* | 0.62* | 0.48* | 0.45* | 0.57* | 0.44* | 0.42* | 0.54* | 0.61* | 0.61* | 0.48* | 0.62* | 0.50* | 0.47* | 0.26* | -0.13* | -0.23* | -0.11* | -0.13 |
| QNW | 0.87* | - | 0.70* | 0.46* | 0.51* | 0.72* | 0.49* | 0.48* | 0.56* | 0.62* | 0.58* | 0.41* | 0.55* | 0.60* | 0.50* | 0.36* | -0.23* | -0.40* | -0.23* | -0.24* |
| REA | 0.71* | 0.76* | - | 0.35* | 0.60* | 0.62* | 0.42* | 0.37* | 0.51* | 0.47* | 0.54* | 0.57* | 0.58* | 0.61* | 0.47* | 0.34* | -0.19* | -0.24* | -0.14* | -0.16* |
| TNP | 0.69* | 0.67* | 0.53* | - | 0.23* | 0.56* | 0.41* | 0.63* | 0.41* | 0.68* | 0.54* | 0.27* | 0.42* | 0.26* | 0.53* | 0.31* | 0.05 | -0.05 | -0.09 | < 0.01 |
| HPB | 0.54* | 0.63* | 0.69* | 0.52* | - | 0.62* | 0.51* | 0.44* | 0.48* | 0.45* | 0.47* | 0.48* | 0.57* | 0.79* | 0.62* | 0.54* | -0.22* | -0.20* | -0.18* | -0.19* |
| TII | 0.69* | 0.78* | 0.65* | 0.69* | 0.70* | - | 0.56* | 0.70* | 0.60* | 0.76* | 0.71* | 0.43* | 0.59* | 0.62* | 0.72* | 0.51* | -0.04 | -0.23* | -0.10* | -0.08 |
| SFD | 0.57* | 0.63* | 0.56* | 0.70* | 0.64* | 0.77* | - | 0.65* | 0.61* | 0.58* | 0.50* | 0.52* | 0.54* | 0.49* | 0.65* | 0.61* | -0.09 | -0.12 | -0.33* | -0.06 |
| IOB | 0.52* | 0.57* | 0.45* | 0.72* | 0.57* | 0.76* | 0.82* | - | 0.60* | 0.78* | 0.67* | 0.42* | 0.51* | 0.46* | 0.75* | 0.57* | 0.03 | -0.07 | -0.16* | -0.02 |
| ETR | 0.57* | 0.61* | 0.61* | 0.56* | 0.54* | 0.60* | 0.61* | 0.63* | - | 0.77* | 0.77* | 0.55* | 0.59* | 0.56* | 0.62* | 0.46* | -0.25* | -0.25* | -0.25* | -0.16* |
| DLC | 0.68* | 0.68* | 0.58* | 0.79* | 0.58* | 0.74* | 0.75* | 0.82* | 0.75* | - | 0.96* | 0.49* | 0.60* | 0.52* | 0.70* | 0.48* | -0.05 | -0.16* | -0.15* | -0.11 |
| APC | 0.64* | 0.63* | 0.58* | 0.67* | 0.55* | 0.68* | 0.63* | 0.68* | 0.78* | 0.85* | - | 0.51* | 0.58* | 0.60* | 0.64* | 0.41* | -0.12 | -0.20* | -0.15* | -0.15* |
| IEF | 0.49* | 0.51* | 0.67* | 0.47* | 0.68* | 0.54* | 0.62* | 0.58* | 0.64* | 0.61* | 0.54* | - | 0.54* | 0.64* | 0.49* | 0.30* | -0.21* | -0.17* | -0.19* | -0.20* |
| TKL | 0.52* | 0.59* | 0.54* | 0.56* | 0.57* | 0.62* | 0.63* | 0.60* | 0.53* | 0.62* | 0.61* | 0.50* | - | 0.61* | 0.69* | 0.48* | -0.11 | -0.15 | -0.08 | -0.05 |
| RAL | 0.56* | 0.60* | 0.70* | 0.51* | 0.84* | 0.63* | 0.60* | 0.56* | 0.58* | 0.60* | 0.55* | 0.79* | 0.62* | - | 0.64* | 0.49* | -0.14* | -0.16* | -0.10 | -0.12 |
| CNL | 0.52* | 0.57* | 0.54* | 0.66* | 0.65* | 0.67* | 0.75* | 0.80* | 0.60* | 0.74* | 0.69* | 0.61* | 0.79* | 0.70* | - | 0.60* | -0.08 | -0.17* | -0.15* | -0.06 |
| NWL | 0.48* | 0.53* | 0.46* | 0.65* | 0.62* | 0.67* | 0.75* | 0.74* | 0.54* | 0.70* | 0.64* | 0.55* | 0.72* | 0.64* | 0.79* | - | -0.02 | -0.11 | -0.15* | -0.02 |
| IER | -0.34* | -0.37* | -0.56* | -0.27* | -0.54* | -0.31* | -0.40* | -0.29* | -0.45* | -0.40* | -0.36* | -0.61* | -0.35* | -0.66* | -0.39* | -0.36* | - | 0.93* | 0.85* | 0.90* |
| WHE | -0.70* | -0.85* | -0.75* | -0.55* | -0.61* | -0.73* | -0.62* | -0.52* | -0.54* | -0.63* | -0.58* | -0.54* | -0.58* | -0.59* | -0.55* | -0.50* | 0.50* | - | 0.84* | 0.90* |
| SGN | -0.53* | -0.57* | -0.57* | -0.53* | -0.53* | -0.60* | -0.76* | -0.60* | -0.60* | -0.65* | -0.56* | -0.62* | -0.50* | -0.55* | -0.59* | -0.58* | 0.55* | 0.70* | - | 0.83* |
| SBR | -0.44* | -0.49* | -0.63* | -0.41* | -0.54* | -0.54* | -0.57* | -0.47* | -0.54* | -0.54* | -0.48* | -0.68* | -0.37* | -0.59* | -0.47* | -0.44* | 0.61* | 0.65* | 0.81* | - |

DIM, Dimension; QLW, Quality of work; QNW, Quantity of work; REA, Rule adherence; TNP, Technical performance; HPB, Helpful behaviours; TII, Taking initiative; SFD, Self-development; IOB, Innovative behaviours; ETR, Emotional resilience; DLC, Dealing with complexity; APC, Adapting to crises; IEF, Interpersonal flexibility; TKL, Task-orientated leadership; RAL, Relations-orientated leadership; CNL, Change-orientated leadership; NWL, Network-orientated leadership; IER, Interpersonal rudeness; WHE, Withholding effort; SGN, Stagnation; SBR, Stubborn resistance.

*, $p < 0.05$.

TABLE 3: Measurement invariance tests across ethnic groups on the individual work performance review bifactor models.

| Scale | Model | χ^2 | df | CFI | RMSEA | $\Delta \chi^2$ | Δ CFI | Δ RMSEA |
|------------------------|---------|----------|------|------|-------|-----------------|--------------|----------------|
| In-role performance | Config. | 324.63 | 176 | 0.98 | 0.05 | - | - | - |
| | Metric | 369.69 | 203 | 0.98 | 0.05 | 45.06* | < 0.01 | < 0.01 |
| | Scalar | 399.20 | 214 | 0.98 | 0.05 | 29.51* | < 0.01 | < 0.01 |
| | Strict | 494.60 | 230 | 0.97 | 0.06 | 95.40* | -0.01 | 0.01 |
| Extra-role performance | Config. | 337.80 | 176 | 0.98 | 0.05 | - | - | - |
| | Metric | 395.85 | 203 | 0.98 | 0.05 | 58.05* | < 0.01 | < 0.01 |
| | Scalar | 421.52 | 214 | 0.98 | 0.05 | 25.67* | < 0.01 | < 0.01 |
| | Strict | 454.61 | 230 | 0.97 | 0.06 | 33.09* | < 0.01 | < 0.01 |
| Adaptive performance | Config. | 581.98 | 196 | 0.96 | 0.08 | - | - | - |
| | Metric | 655.70 | 208 | 0.95 | 0.08 | 73.72* | -0.01 | < 0.01 |
| | Scalar | 714.94 | 220 | 0.95 | 0.08 | 59.24* | < 0.01 | < 0.01 |
| | Strict | 787.80 | 236 | 0.94 | 0.08 | 72.86* | -0.01 | < 0.01 |
| Leadership performance | Config. | 425.96 | 176 | 0.97 | 0.07 | - | - | - |
| | Metric | 475.36 | 203 | 0.97 | 0.07 | 49.40* | < 0.01 | < 0.01 |
| | Scalar | 498.32 | 214 | 0.97 | 0.06 | 22.96* | < 0.01 | < 0.01 |
| | Strict | 556.16 | 230 | 0.97 | 0.07 | 57.84* | -0.01 | < 0.01 |
| Counter performance | Config. | 247.39 | 176 | 0.98 | 0.05 | - | - | - |
| | Metric | 318.24 | 203 | 0.97 | 0.05 | 70.85* | -0.01 | 0.01 |
| | Scalar | 343.77 | 214 | 0.97 | 0.06 | 25.53* | < 0.01 | < 0.01 |
| | Strict | 547.61 | 230 | 0.91 | 0.08 | 203.84* | -0.06 | 0.03 |

CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation.

*, $p < 0.05$.

The upper limits of the inter-factor correlations were, therefore, not inspected. Instead, greater attention was paid to the differences in inter-factor correlations between self- and managerial-ratings.

The inter-factor correlations were, on average, higher for managerial-ratings ($|\phi| = -0.85$ to 0.87 ; $M = 0.60$), compared to self-ratings ($|\phi| = -0.40$ to 0.96 ; $M = 0.42$). Managers appeared to observe a more consistent pattern in performance across the narrow performance dimensions, whereas employees tended to discriminate more between the dimensions when rating their own performance. The difference in inter-factor correlations in self-ratings ($|\phi| = -0.40$ to 0.05 ; $M = 0.14$) versus managerial-ratings ($|\phi| = -0.85$ to -0.27 ; $M = 0.53$) appeared particularly salient when considering the smaller and nonsignificant inter-factor correlations between the narrow dimensions of *counterproductive performance* and other performance dimensions. It could be that managers form more overall impressions, which could be influenced by the halo or horn effect, instead of discerning more specific behaviours, as would be the case when employees rate themselves (Dalal et al., 2009; Spector et al., 2010; Vandenberg et al., 2005). It is also possible that employees may want to minimise their managers' perceptions of their degree of counterproductive performance, thereby, engaging in leniency bias, because of the ego-threatening nature of self-ratings on *counterproductive performance* (Aguinis, 2019; Spector et al., 2010). Consequently, employees might unconsciously or deliberately uncouple self-ratings on *counterproductive performance* from other performance dimensions.

Table 3 presents the structural invariance of each of the five broad performance dimensions. Four of the five models were specified to be hierarchical (bifactor models) in nature, for example, an orthogonal general factor was specified for *in-role performance*, along with orthogonal group factors for

quality of work, quantity of work, rule adherence, and technical performance. *Adaptive performance*, in contrast, was specified to be an oblique lower-order model. The narrow dimensions of *adaptive performance*, especially *dealing with complexity* and *adapting to crises*, had higher inter-factor correlations for self-ratings when compared to managerial ratings, as evident in Table 2. The researchers were unable to identify the multi-group bifactor model when specifying configural invariance, and, therefore, opted for a more parsimonious model, namely, an oblique lower-order model, which allowed the narrow dimensions in *adaptive performance* to correlate freely. Even though the general factor of *adaptive performance* was not specified, the inter-factor correlations in the model for self-ratings were large enough ($|\phi| = 0.49$ to 0.96 ; $M = 0.68$) to suggest the presence of a strong general factor. Note that the hierarchical structure of *adaptive performance* for managerial ratings was inspected by Van Lill and Taylor (2022).

All the models, except for the strict invariance of *counterproductive performance*, conveyed sufficient structural equivalence. The factor covariances on *counterproductive performance* appeared to be much higher for self-ratings ($|\phi| = 0.83$ to 0.93 ; $M = 0.88$), compared to managerial ratings ($|\phi| = 0.50$ to 0.81 ; $M = 0.66$), and suggested that employees may use a smaller range of the construct continua when rating themselves (Scullen et al., 2003; Vandenberg & Lance, 2000). Employees may have consistently given themselves low ratings across items and dimensions of *counterproductive performance* because of the ego-threatening nature of the content evaluated, which served as clarification of the lower strict invariance.

The researchers considered the existing evidence in support of the measurement invariance of the five broad performance dimensions sufficient to prove that employees (when rating

TABLE 4: *T*-test results, narrow performance scale means, standard deviations, and reliability coefficients per rating group.

| Scale | Group | Σ | SD | Trim | α | ω | <i>t</i> | <i>df</i> | <i>d</i> |
|---------------------------|---------|----------|-------|-------|----------|----------|----------|-----------|----------|
| In-role performance | Manager | 66.85 | 10.82 | 68.12 | 0.95 | 0.87 | 5.81 | 449.45 | 0.31* |
| | Self | 71.69 | 6.39 | 72.28 | 0.88 | 0.74 | | | |
| Quality of work | Manager | 16.50 | 2.94 | 16.77 | 0.87 | 0.87 | 4.03 | 484.13 | 0.21* |
| | Self | 17.55 | 2.03 | 17.72 | 0.68 | 0.69 | | | |
| Quantity of work | Manager | 16.92 | 3.14 | 17.38 | 0.93 | 0.93 | 5.27 | 503.72 | 0.29* |
| | Self | 18.23 | 1.89 | 18.49 | 0.83 | 0.83 | | | |
| Rule adherence | Manager | 17.58 | 2.57 | 17.97 | 0.83 | 0.84 | 6.48 | 493.19 | 0.30* |
| | Self | 18.68 | 1.62 | 18.95 | 0.71 | 0.72 | | | |
| Technical performance | Manager | 15.85 | 3.99 | 16.31 | 0.95 | 0.96 | 3.87 | 455.82 | 0.22* |
| | Self | 17.23 | 2.70 | 17.57 | 0.89 | 0.91 | | | |
| Extra-role performance | Manager | 61.81 | 12.90 | 62.78 | 0.96 | 0.88 | 9.50 | 429.94 | 0.50* |
| | Self | 70.15 | 7.14 | 70.68 | 0.91 | 0.79 | | | |
| Helpful behaviours | Manager | 16.87 | 3.11 | 17.26 | 0.93 | 0.93 | 6.37 | 420.03 | 0.36* |
| | Self | 18.35 | 1.90 | 18.64 | 0.81 | 0.82 | | | |
| Taking initiative | Manager | 15.93 | 3.88 | 16.43 | 0.94 | 0.94 | 6.43 | 416.11 | 0.39* |
| | Self | 18.02 | 2.00 | 18.28 | 0.79 | 0.79 | | | |
| Self-development | Manager | 15.25 | 3.90 | 15.61 | 0.93 | 0.93 | 8.54 | 482.29 | 0.44* |
| | Self | 17.52 | 2.38 | 17.80 | 0.85 | 0.85 | | | |
| Innovative behaviours | Manager | 13.76 | 3.98 | 13.89 | 0.95 | 0.95 | 9.94 | 494.47 | 0.45* |
| | Self | 16.27 | 2.78 | 16.46 | 0.88 | 0.88 | | | |
| Adaptive performance | Manager | 61.22 | 12.20 | 62.07 | 0.96 | 0.91 | 9.74 | 464.86 | 0.51* |
| | Self | 69.55 | 7.39 | 70.04 | 0.93 | 0.88 | | | |
| Emotional resilience | Manager | 15.44 | 3.48 | 15.73 | 0.93 | 0.93 | 6.41 | 493.56 | 0.38* |
| | Self | 17.14 | 2.40 | 17.37 | 0.85 | 0.85 | | | |
| Dealing with complexity | Manager | 14.73 | 3.66 | 14.97 | 0.94 | 0.94 | 10.68 | 473.14 | 0.49* |
| | Self | 17.23 | 2.28 | 17.45 | 0.85 | 0.85 | | | |
| Adapting to crises | Manager | 14.90 | 3.64 | 15.13 | 0.95 | 0.95 | 9.47 | 475.33 | 0.43* |
| | Self | 16.99 | 2.32 | 17.18 | 0.84 | 0.85 | | | |
| Interpersonal flexibility | Manager | 16.15 | 3.36 | 16.54 | 0.93 | 0.93 | 8.68 | 480.36 | 0.45* |
| | Self | 18.19 | 1.89 | 18.41 | 0.83 | 0.83 | | | |
| Leadership performance | Manager | 59.57 | 13.56 | 60.36 | 0.97 | 0.91 | 11.37 | 453.74 | 0.58* |
| | Self | 69.96 | 7.84 | 70.60 | 0.93 | 0.82 | | | |
| Task leadership | Manager | 14.67 | 4.08 | 14.97 | 0.97 | 0.97 | 9.36 | 482.57 | 0.41* |
| | Self | 17.14 | 2.60 | 17.44 | 0.90 | 0.90 | | | |
| Relation leadership | Manager | 16.42 | 3.17 | 16.73 | 0.92 | 0.92 | 10.43 | 419.71 | 0.58* |
| | Self | 18.61 | 1.70 | 18.85 | 0.81 | 0.82 | | | |
| Change leadership | Manager | 13.98 | 3.94 | 14.12 | 0.95 | 0.95 | 12.09 | 493.49 | 0.55* |
| | Self | 17.06 | 2.63 | 17.34 | 0.91 | 0.91 | | | |
| Network leadership | Manager | 14.50 | 4.13 | 14.76 | 0.93 | 0.93 | 9.56 | 464.37 | 0.47* |
| | Self | 17.18 | 2.60 | 17.47 | 0.85 | 0.87 | | | |
| Counter performance | Manager | 23.91 | 9.38 | 22.23 | 0.94 | 0.84 | 7.00 | 336.45 | 0.44* |
| | Self | 19.47 | 6.64 | 18.25 | 0.94 | 0.91 | | | |
| Interpersonal rudeness | Manager | 5.93 | 2.74 | 5.35 | 0.89 | 0.89 | 2.87 | 417.13 | 0.16* |
| | Self | 5.16 | 1.92 | 4.78 | 0.80 | 0.79 | | | |
| Withholding effort | Manager | 5.95 | 2.72 | 5.41 | 0.88 | 0.89 | 5.29 | 320.83 | 0.37* |
| | Self | 4.89 | 1.81 | 4.51 | 0.84 | 0.84 | | | |
| Stagnation | Manager | 6.13 | 3.04 | 5.52 | 0.88 | 0.89 | 6.82 | 310.66 | 0.35* |
| | Self | 4.79 | 1.94 | 4.35 | 0.88 | 0.88 | | | |
| Stubborn resistance | Manager | 5.91 | 2.86 | 5.32 | 0.91 | 0.91 | 6.26 | 304.57 | 0.46* |
| | Self | 4.63 | 1.77 | 4.22 | 0.86 | 0.86 | | | |

Note: Values of 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes (Mair & Wilcox, 2020).

Σ , mean of total scores; Trim, 20% trimmed mean of total scores; *t*, Yuen's test of trimmed means; *df*, degrees of freedom; *d*, Explanatory measure of effect size; ω , omega hierarchical reliability for broad dimensions and omega conditional on other factors for narrow dimensions; SD, standard deviation.

*, $p < 0.05$.

themselves) and managers (when rating their subordinates) have a similar conceptual understanding of the performance dimensions in question. Hypothesis 1 was, therefore, supported, which enabled further inspection of the mean group differences in the performance dimensions between self- and managerial-ratings.

Table 4 presents the mean group differences in self- and managerial-ratings for each of the 20 narrow performance

dimensions used in the present study. The mean group differences were also calculated for the broad five performance dimensions.

The mean differences on the narrow performance dimensions, reported in Table 4, on average ($|d| = 0.16$ to 0.58 ; $M = 0.39$), suggest that employees tend to give themselves more lenient ratings (in the moderate to large range), compared to managerial ratings. The present research

was, therefore, able to replicate the meta-analytical findings of Heidemeier and Moser (2009) and Conway and Huffcutt (1997), although the average size of the mean group differences was slightly higher than 0.32 reported by Heidemeier and Moser (2009). Preliminary evidence, therefore, supported Hypothesis 2. It further appeared, like the findings reported by Heidemeier and Moser (2009), that the mean group differences for *in-role performance* ($d = 0.31$), also referred to as 'task performance', were lower when compared to the mean group differences for the other four broad dimensions ($|d| = 0.44$ to 0.58 ; $M = 0.51$). Employees are thus likely to receive continuous feedback on explicit expectations, as is the case with the narrow dimensions of *in-role performance*, and may, therefore, hold more calibrated judgements of their self-ratings when compared with those of managers (Heidemeier & Moser, 2009).

Discussion

Outline of the results

Drawing comparisons between self-ratings and other ratings has become a crucial aspect of performance development. However, continued research is required to determine whether employees (when rating themselves) and other rating sources have a similar conceptual understanding of the construct under investigation. Furthermore, it is also important to determine whether there are any significant differences in the magnitude of ratings that employees give themselves versus the ratings provided by others.

The present study was a first of its kind in South Africa to investigate the measurement invariance and mean group differences in broad and narrow generic performance dimensions among self-ratings versus managerial ratings. Preliminary evidence suggests that employees have a similar understanding to managers when rating themselves on the dimensions identified in the IWPR. Furthermore, employees tended to consistently give themselves higher ratings compared to the ratings given by managers, mostly in the moderate-to-large range. It is likely that employees, whether consciously or unconsciously, rely on a leniency heuristic when rating their own performance. It is worth mentioning that this self-deceiving bias might be lower when performance evaluations are completed for research purposes and, therefore, kept anonymous. It is also possible that leniency bias will be higher when performance evaluations are completed for administrative (rewards or promotions) purposes (Scullen et al., 2003).

Practical implications

Managers

At a base level, there seems to be an overlap between how managers and employees conceptualise performance dimensions on the IWPR. This serves as evidence that a common performance language is used when decisions about performance need to be made or performance feedback must be provided. However, self-ratings of performance in South Africa appear to be overly lenient, and it might be prudent to give managerial-ratings more

weight, or ignore self-ratings, when high-stakes decisions about employees must be made, such as rewards or promotions. That said, employees might still have a more nuanced understanding of their own behaviour at work, as evident in the lower inter-factor correlations displayed for self-ratings in this study, which could give managers a more in-depth view of areas for performance development. It might, therefore, be meaningful to report self-ratings and managerial-ratings separately – and visually compare differences – to clarify where possible discrepancies exist. Self-ratings can, therefore, serve as an anchor in performance feedback, to determine the degree to which self-perceptions – and subsequent behaviour – need to be adjusted in order to meet performance expectations. In contrast, the aggregation of self- and managerial-ratings into one overall score might be a less helpful strategy.

Task-based feedback, based on *in-role performance*, might come as less of a surprise when performance feedback is provided by managers, given the shared understanding of expectations on the relevant performance dimension. By contrast, discrepancies in ratings might be particularly high for *extra-role*, *adaptive*, *leadership*, and *counterproductive performance*. Feedback on the latter four dimensions, especially when high discrepancies between self- and managerial ratings exist, could be taken as a personal attack, and could rapidly escalate to unhelpful conflict. Therefore, managers are encouraged to continuously observe everyday behaviours and provide feedback at frequent intervals, ensuring that such feedback remains work-relevant and does not come across as personal attacks (Aguinis, 2019). An industrial psychologist could also be asked to act as a mediator in difficult conversations, ensuring that the performance conversation remains objective, work-relevant, and constructive.

Employees

It is important that the rationale behind the goal of self- and managerial ratings is clearly communicated to employees, increasing their perception of the fairness of the process. When treated with respect and given information timely, employees might be more likely to hold more positive expectations of the process, experience a sense of control, and, therefore, commit to the performance objectives of the organisation (Van Lill et al., 2020).

Industrial psychologists

Talent analytics require accurate criterion data. Therefore, it might be more prudent for industrial psychologists to use managerial ratings when determining the utility of different selection procedures or development initiatives. Managerial ratings also come with the added benefit that method bias is, to some extent, reduced when another rating source is used to conduct local concurrent validation studies (Podsakoff et al., 2012). Industrial psychologists, as custodians of people data, also have an important responsibility to educate organisational stakeholders on why self- and managerial ratings should be unteased and

differentially applied in selection or development contexts within the workplace. Such explanations might also increase the perceived fairness of the performance evaluation process. Finally, given the clear discrepancies between self- and managerial ratings, it might be useful for industrial psychologists to develop different norms when reporting self- and managerial scores for performance feedback purposes. In doing so, the biases of different rating sources could be controlled to some extent.

Limitations and recommendations

The present study focussed on only one alternative rating source to self-ratings, namely, managerial ratings. Future studies could also inspect the measurement invariance and mean group differences of self-ratings versus peer, self-ratings versus subordinate, or self-rating versus customer ratings. This will enable a more holistic view of the way in which scores outside of self-ratings could be aggregated to provide a comprehensive 360-degree view of employees' reputations at work.

This study relied on different cohorts' self- and managerial-ratings. Consequently, it was not possible to calculate the inter-rater reliabilities of the performance dimensions measured across different rating sources. Inspections of the inter-rater reliability could provide meaningful guidelines according to which rating scores may be meaningfully clustered before rating aggregations are made.

Finally, the explicit purpose of the study was to use the data for development purposes. Future studies could consider investigating differences in rating sources according to use for different purposes, such as research, promotion, or remuneration. The current researchers expect larger mean group differences when the IWPR is used for administrative purposes, such as promotion or remuneration.

Conclusion

The use of 360-degree evaluations is an increasingly popular method to rate employees' performance. However, the meaningful application of 360-degree feedback in decision-making requires a purposeful investigation of a shared conceptual understanding across rating sources, as well as population-wide rating discrepancies. The present study aimed to investigate the measurement invariance and mean group differences of several broad and narrow dimensions of performance on the IWPR across self- and managerial-ratings. The evidence suggests a sufficient degree of overlap between the factor structure of the measurement to make meaningful comparisons. Furthermore, employees appear to be more lenient in their self-ratings when compared to managerial ratings.

Acknowledgements

We would like to thank our colleagues at JVR Africa Group for enriching our understanding, through conversation, on

the differences in self- versus managerial-ratings of generic performance.

Competing interests

Both authors are employees of JVR Africa Group, which is the company for which this instrument was developed.

Authors' contributions

X.v.L. and G.v.d.M. developed the conceptual framework and devised the method. X.v.L. analysed the data.

Funding information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

Coefficients based on the multi-group confirmatory factor analysis (CFA) are available on request.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors, and the publisher.

References

- Aguinis, H. (2019). *Performance management* (4th ed.). Chicago Business Press.
- Aguinis, H., & Edwards, J.R. (2014). Methodological wishes for the next decade and how to make wishes come true. *Journal of Management Studies*, 51(1), 143–174. <https://doi.org/10.1111/joms.12058>
- Atwater, L.E., & Yammarino, F.J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45(1), 141–164. <https://doi.org/10.1111/j.1744-6570.1992.tb00848.x>
- Bandalos, D.L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling*, 21(1), 102–116. <https://doi.org/10.1080/10705511.2014.859510>
- Borman, W.C., & Motowidlo, S.J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10(2), 99–109. https://doi.org/10.1207/s15327043hup1002_3
- Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Brown, T.A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Browne, M.W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- Campbell, J.P., & Wiernik, B.M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 47–74. <https://doi.org/10.1146/annurev-orgpsych-032414-111427>
- Carpini, J.A., Parker, S.K., & Griffin, M.A. (2017). A look back and a leap forward: A review and synthesis of the individual work performance literature. *Academy of Management Annals*, 11(2), 825–885. <https://doi.org/10.5465/annals.2015.0151>
- Casper, W.C., Edwards, B.D., Wallace, J.C., Landis, R.S., & Fife, D.A. (2020). Selecting response anchors with equal intervals for summated rating scales. *Journal of Applied Psychology*, 105(4), 390–409. <https://doi.org/10.1037/apl0000444>
- Conway, J.M., & Huffcutt, A.I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10(4), 331–360. https://doi.org/10.1207/s15327043hup1004_2
- Dalal, R.S., Lam, H., Weiss, H.M., Welch, E.R., & Hulin, C.L. (2009). A within-person approach to work behavior and performance: Concurrent and lagged citizenship-counter productivity associations, and dynamic relationships with affect and overall job performance. *Academy of Management Journal*, 52(5), 1051–1066. <https://doi.org/10.5465/amj.2009.44636148>

- Facteau, J.D., & Craig, B.S. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, 86(2), 215–227. <https://doi.org/10.1037/0021-9010.86.2.215>
- Harari, M.B., & Viswesvaran, C. (2018). Individual job performance. In D.S. Ones, N. Anderson, C. Viswesvaran, & H.K. Sinangil (Eds.), *The Sage handbook of industrial, work, and organizational psychology: Personnel psychology and employee performance* (pp. 55–72). Sage.
- Heidemeier, H., & Moser, K. (2009). Self–other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology*, 94(2), 353–370. <https://doi.org/10.1037/0021-9010.94.2.353>
- Hogan, R., & Sherman, R.A. (2020). Personality theory and the nature of human nature. *Personality and Individual Differences*, 152(2020), 1–5. <https://doi.org/10.1016/j.paid.2019.109561>
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, 52(2), 464–488. <https://doi.org/10.3758/s13428-019-01246-w>
- Mair, P., & Wilcox, R. (2021). A collection of robust statistical methods. Package WRS2. Retrieved from <https://cran.r-project.org/web/packages/WRS2/WRS2.pdf>
- Marcus, B., Taylor, O.A., Hastings, S.E., Sturm, A., & Weigelt, O. (2016). The structure of counterproductive work behavior: A review, a structural meta-analysis, and a primary study. *Journal of Management*, 42(1), 203–233. <https://doi.org/10.1177/0149206313503019>
- Motowidlo, S.J., & Van Scotter, J.R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79(4), 475–480. <https://doi.org/10.1037/0021-9010.79.4.475>
- Myburgh, H.M. (2013). *The development and evaluation of a generic individual non-managerial performance measure*. Unpublished master's dissertation, Stellenbosch University, Stellenbosch. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.985.2008&rep=rep1&type=pdf>
- Organ, D.W. (1997). Organizational citizenship behavior: It's construct clean-up time. *Human Performance*, 10(2), 85–97. https://doi.org/10.1207/s15327043hup1002_2
- Podsakoff, P.M., MacKenzie, S.B., & Podsakoff, N.P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63(1), 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Pulakos, E.D., Arad, S., Donovan, M.A., & Plamondon, K.E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, 85(4), 612–624. <https://doi.org/10.1037/0021-9010.85.4.612>
- Pulakos, E.D., Mueller-Hanson, R., & Arad, S. (2019). The evolution of performance management: Searching for value. *Annual Review of Organizational Psychology and Organizational Behavior*, 6(1), 1–23. <https://doi.org/10.1146/annurev-orgpsych-012218-015009>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from <https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf>
- Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rosseel, Y., Jorgensen, T., & Rockwood, N. (2022). *Latent variable analysis*. Retrieved from <https://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
- Rousseau, D.M. (Ed.). (2012). *Envisioning evidence-based management*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199763986.013.0001>
- Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C.C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.
- Schepers, J.M. (2008). The construction and evaluation of a generic work performance questionnaire for use with administrative and operational staff. *SA Journal of Industrial Psychology*, 34(1), 10–22. <https://doi.org/10.4102/sajip.v34i1.414>
- Scullen, S.E., Mount, M.K., & Judge, T.A. (2003). Evidence of the construct validity of developmental ratings of managerial performance. *Journal of Applied Psychology*, 88(1), 50–66. <https://doi.org/10.1037/0021-9010.88.1.50>
- Spector, P.E. (2019). Do not cross me: Optimizing the use of cross-sectional designs. *Journal of Business and Psychology*, 34(2), 125–137. <https://doi.org/10.1007/s10869-018-09613-8>
- Spector, P.E., Bauer, J.A., & Fox, S. (2010). Measurement artifacts in the assessment of counterproductive work behavior and organizational citizenship behavior: Do we know what we think we know? *Journal of Applied Psychology*, 95(4), 781–790. <https://doi.org/10.1037/a0019477>
- Spector, P.E., & Fox, S. (2005). The stressor-emotion model of counterproductive work behavior. In P.E. Spector & S. Fox (Eds.), *Counterproductive work behavior* (pp. 151–174). American Psychological Association.
- Van der Vaart, L. (2021). The performance measurement conundrum: Construct validity of the Individual Work Performance Questionnaire in South Africa. *South African Journal of Economic and Management Sciences*, 24(1), 1–11. <https://doi.org/10.4102/sajems.v24i1.3581>
- Van Lill, X., Roodt, G., & De Bruin, G.P. (2020). The relationship between managers' goal-setting styles and subordinates' goal commitment. *South African Journal of Economic and Management Sciences*, 23(1), 1–11. <https://doi.org/10.4102/sajems.v23i1.3601>
- Van Lill, X., & Taylor, N. (2022). The validity of five broad generic dimensions of performance in South Africa. *South African Journal of Human Resource Management*, 20, 1–15. <https://doi.org/10.4102/sajhrm.v20i0.1844>
- Van Lill, X., & Van der Vaart, L. (2022). *The validity of a general factor of individual work performance in the South African context*. [Manuscript submitted for publication].
- Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–69. <https://doi.org/10.1177/109442810031002>
- Vandenberg, R.J., Lance, C.E., & Taylor, S.C. (2005). A latent variable approach to rating source equivalence: Who should provide ratings on organizational citizenship behavior dimensions? In D.L. Turnipseed (Ed.), *Handbook of organizational citizenship behavior* (pp. 109–141). Nova Science.
- Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, 90(1), 108–131. <https://doi.org/10.1037/0021-9010.90.1.108>
- Yuan, K.H., & Bentler, P.M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology*, 51(2), 289–309. <https://doi.org/10.1111/j.2044-8317.1998.tb00682.x>
- Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives*, 26(4), 66–85. <https://doi.org/10.5465/amp.2012.0088>