TECHNICAL BRIEF

Exploring Gender Differences on the Hogan 360

Authored by:
Dr Xander van Lill, Organisational Psychologist, PBC
Dr Lynne Cruickshank, Head of Research, PBC

Introduction

360 feedback, provided in the appropriate context, can help leaders maintain or increase the effectiveness with which they help their teams achieve organisational goals (Smither et al., 2005). However, the impact of 360 feedback depends on the quality of the tool used, the user's knowledge about evidence-based practice around behavioural change, and the user's skill in catalysing a leader's sustained motivation to achieve outcomes. 360 feedback could demotivate leaders if invalid and biased measurements are used (Nowack and Mashihi, 2012). Evidence on the reliability and validity of the Hogan 360 overall is presented in the Hogan 360 Technical Manual (Peter Berry Consultancy, 2022). This technical brief summarises supplementary analyses examining measurement equivalence and mean group differences by gender. The analyses provide information on whether biases may impact how raters respond to items when evaluating men and women on the Hogan 360, and whether any practically meaningful mean group differences are observed between men and women in others' ratings.

A comprehensive meta-analysis by Roth et al. (2012) examined gender differences in job performance ratings across various field studies. Their findings indicated that men received slightly higher ratings than women, although the effect size was small, suggesting only a modest, context-dependent disparity. The authors noted that job type and organisational culture may help explain these differences. More recently, a meta-analysis by Mackey et al. (2017) explored the effects of gender proportionality in organisations (i.e., whether women's job performance ratings change depending on the proportion of women in the organisation) and found that women tended to demonstrate slightly higher job performance than men. Importantly, this performance advantage did not vary as a function of the proportion of women in the organisation, challenging tokenism-based theories that predict performance disadvantages for gender minorities. Drawing on these findings, we did not anticipate substantial gender-related differences in Hogan 360 performance outcomes. Nonetheless, the current analyses investigates the potential influence of gender on ratings to support the Hogan 360's ongoing fairness and interpretive accuracy.

Method

Sample

Data was analysed for a sample of 44,212 subjects who completed the Hogan 360 between 1st January 2013, and 31st December 2023. Most of the sample was men (n = 23,236, 53%), followed by women (n = 14,830, 34%). The rest of the sample were unspecified in terms of their self-identified gender (n = 6,146, 13%)

Instrument

The Hogan 360, developed by Peter Berry Consultancy, is a comprehensive multi-rater survey designed to gather leadership feedback from key stakeholder groups, such as managers, peers, and direct reports. This tool assesses four main domains and 14 underlying competency subthemes, providing a holistic view of leadership effectiveness.

In its current format, the Hogan 360 features:

- 50 scaled items rated on a 7-point scale, with 1 meaning 'Does not describe this person at all' and 7 meaning 'Describes this person exactly'. These items are aligned with the four domains of the Hogan 360 Leadership Model and their respective sub-themes.
- Ranked items that highlight the subject's top strengths and areas for improvement. Raters choose the subject's top four strengths and opportunities to improve from 30 items, with the top selection given the highest weight of 4, followed by weights of 3, 2, and 1.
- Three open-ended questions focused on identifying strengths, opportunities, and overused strengths.

This research focuses on data from the scaled items.

Figure 1

The Hogan 360 Leadership Model



Analysis

Measurement invariance refers to the statistical property that ensures a measurement tool (e.g., a survey, test, or assessment) functions equivalently across different groups or contexts. This means that the instrument measures the same underlying construct in the same way, regardless of factors such as cultural background, gender, or language. Establishing measurement invariance is critical for making valid comparisons between groups, as it confirms that the observed differences are due to the construct being measured and not due to bias in the measurement tool (Brown, 2015).

In this technical brief, we followed the sequence recommended by Vandenberg and Lance (2000) to test measurement invariance across multiple groups. We determined the equivalence of an oblique lower-order model for the Hogan 360 domains. Measurement invariance was assessed by comparing the fit statistics of a multigroup confirmatory factor model for configural invariance (equality of the factor structure across groups), metric invariance (similar factor loadings across groups), scalar invariance (similar intercepts across groups), and strict invariance (similar residual variances), using Version 0.6-19 of the lavaan package (Rosseel, 2012; Rosseel et al., 2024) in R (R Core Team, 2024).

Mardia's multivariate skewness (142367.90; $p < .01$) and kurtosis (777.80; $p < .01$) coefficients indicated that the data had a non-normal multivariate distribution. Given the use of rating scales with seven numerical categories (Rhemtulla et al., 2012) and the violation of multivariate normality (Satorra & Bentler, 1994; Yuan & Bentler, 1998), robust maximum likelihood (MLM) estimation was deemed appropriate (Bandalos, 2014). The model-data fit of the Confirmatory Factor Analysis (CFA) was evaluated using the comparative fit index (CFI), the Tucker–Lewis Index (TLI), the standardized root mean-square residual (SRMR), and the root mean square error of approximation (RMSEA) (Brown, 2015; Hu & Bentler, 1999). The fit was considered suitable if the RMSEA and SRMR were $\leq .08$ (Brown, 2006; Browne & Cudeck, 1992) and the CFI and TLI were $\geq .90$ (Brown, 2006; Hu & Bentler, 1999). Brown (2015) argues that even if goodness-of-fit measures display a marginally good fit to the data (CFI and TLI in the range of .90 to .95), models may still be considered to display acceptable fit if other indices (SRMR and RMSEA) are within the acceptable range.

After inspecting bias with the test of invariance, robust t-tests were conducted on the 14 subthemes of the Hogan 360 using Version 1.1-6 of the WRS2 package in R (Mair & Wilcox, 2020, 2024) to inspect the mean group differences between men and women.

Results

The inter-factor correlations for men and women are reported in **Table 1**. The fit statistics for the oblique lower-order confirmatory factor model of the 14 domains for men (2 [df] = 16705.05 [1084]; CFI = .90; TLI = .89; SRMR = .05; RMSEA = .07 [.06; .07]) and women (2 [df] = 10921.91 [1084]; CFI = .90; TLI = .89; SRMR = .05; RMSEA = .06 [.06; .07]) were satisfactory (Brown, 2015; Hu & Bentler, 1999). Both groups' Tucker-Lewis Index (TLI) is .01 below the recommended cutoff. However, model complexity could influence the TLI due to the number of mathematical parameters specified (Brown, 2015). Viewed in conjunction with the CFI, SRMR, and RMSEA, we conceive that the baseline models for both genders still display a reasonably good fit to proceed with further tests for measurement invariance.

Table 1

Latent Factor Correlations From the 14-factor Confirmatory Factor Analysis for Men (Under the Diagonal) and Women (Above the Diagonal)

Sub-Theme	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Integrity	-	.91	.74	.91	.85	.70	.66	.64	.72	.61	.53	.80	.60	.62
2. Resilience	.92	-	.83	.94	.88	.77	.68	.71	.74	.64	.56	.85	.67	.69
3. Communication	.76	.82	-	.87	.97	.97	.85	.91	.86	.79	.86	.92	.91	.89
4. People Skills	.91	.94	.88	-	.97	.80	.71	.71	.75	.70	.57	.92	.66	.69
5. Team Player	.86	.87	.98	.97	-	.91	.81	.83	.84	.77	.73	.97	.81	.82
6. Engaging	.68	.74	.95	.78	.89	-	.87	.87	.88	.80	.84	.89	.90	.92
7. Capability	.67	.66	.82	.68	.79	.85	-	.83	.91	.73	.72	.74	.78	.80
8. Efficiency	.66	.72	.92	.72	.83	.88	.80	-	.89	.74	.80	.79	.87	.82
9. Results	.73	.75	.88	.75	.84	.88	.88	.89	-	.73	.74	.77	.81	.82
10. Customer	.61	.64	.80	.69	.78	.81	.68	.75	.73	-	.69	.74	.75	.78
11. Accountability	.54	.55	.85	.57	.72	.80	.67	.83	.74	.68	-	.71	.84	.77
12. Motivation	.82	.86	.92	.93	.97	.87	.71	.80	.78	.74	.69	-	.80	.81
13. Strategy	.59	.65	.91	.67	.81	.90	.74	.88	.81	.76	.84	.80	-	.94
14. Innovation	.59	.64	.87	.66	.80	.90	.78	.82	.82	.78	.76	.77	.94	-

Note. All the correlations were significant at the .05 level. The results are based on 5069 raters for men and 3205 for women. A smaller subset of data was analysed due to the need for complete item responses from raters. Rater categories for men and women included direct reports, peers, managers, and ther stakeholders.

The pattern of inter-factor correlations reported between men and women was the same, suggesting a fair degree of similarity in the underlying structure of the data, regardless of subjects' gender. We then proceeded to explore the invariance of the model across raters that evaluated men versus women, of which the findings are reported in **Table 2.**

Table 2

Measurement Invariance Test Across Gender Groups on the Hogan 360 Lower-Order Model

Model	$\chi 2$	*df*	CFI	RMSEA	Δ $\chi 2$	Δ CFI	Δ RMSEA
Config	27668.50	2168	.90	.07	-	-	-
Metric	27752.19	2204	.90	.06	36	< .01	< .01
Scalar	28126.34	2240	.90	.06	36	< .01	< .01
Strict	28375.25	2290	.90	.06	50	< .01	< .01

Note. All the correlations were significant at the .05 level. The results are based on 5069 raters for men and 3205 for women.
The following is a brief explanation of important coefficients reported.

- CFI = Compares how well your model fits the data against a baseline model where variables are uncorrelated. It ranges from 0 to 1, with higher values showing a better fit. A CFI below 0.90 suggests the model might not fit well.
- Δ CFI = When testing measurement invariance, we investigate changes in CFI between models with and without constraints. Constraints are conditions applied to ensure specific parts of the model, such as factor loadings, intercepts, or error variances, are equal across groups. These constraints test whether the model behaves the same way for different groups. If the change in CFI (ΔCFI) is less than 0.01 when constraints are added, it suggests the constraints do not significantly harm the model fit, supporting the equivalence of the measure across groups.
- RMSEA = Shows how well the model fits the data by measuring the difference between the model's predictions and the actual data, adjusted for its complexity (number of mathematic elements specified). Values over 0.08 suggest a poor fit.
- ΔRMSEA = When testing measurement invariance, we check changes in RMSEA between models with and without constraints. Like the difference in CFI, these constraints test whether the model behaves the same way for different groups. If the change in RMSEA (ΔRMSEA) is less than 0.015 when added, it suggests the constraints do not significantly harm the model fit, supporting measurement invariance.

The ΔCFI (< 0.01) and ΔRMSEA (< 0.015) provided sufficient evidence for measurement invariance in factor loadings, intercepts, and residuals when comparing raters who evaluated men compared to women (Chen, 2007; Cheung & Rensvold, 2002). Raters, therefore, have the same underlying assumptions about the relation of the items to the Hogan 360 sub-themes, regardless of whether they are rating men or women. Given that we were able to establish the measurement invariance of the underlying factor structure when raters evaluate men and women, we proceeded to inspect mean group differences between the ratings that subjects obtained on the Hogan 360, which are reported in **Table 3.**

Table 3

T-test Results, Sub-Theme Means, Standard Deviations, and Reliability Coefficients for Men and Women

Model	Group	Mean	*SD*	Alpha	Omega	*t*	*df*	*d*
Integrity	Men	5.83	.54	.86	.86	25.49	20457.24	.21 (.19; .23)
	Women	5.98	.49	.86	.86			
Resilience	Men	5.67	.58	.89	.89	18.53	19019.01	.14 (.12; .15)
	Women	5.77	.56	.88	.89			
Communication	Men	5.52	.57	.80	.80	16.08	19275.15	.14 (.11; .15)
	Women	5.62	.52	.80	.80			
People Skills	Men	5.51	.64	.92	.92	31.91	20207.59	.24 (.23; .26)
	Women	5.71	.61	.92	.92			
Team Player	Men	5.42	.61	.89	.89	26.28	20354.45	.20 (.18; .21)
	Women	5.57	.57	.89	.89			
Engaging	Men	5.65	.58	.87	.87	20.51	20512.47	.16 (.14; .17)
	Women	5.77	.52	.86	.85			
Capability	Men	6.01	.49	.84	.85	12.96	20677.01	.10 (.08; .11)
	Women	6.08	.45	.82	.83			
Efficiency	Men	5.52	.59	.89	.89	27.14	20417.44	.21 (.19; .22)
	Women	5.67	.54	.89	.89			
Results	Men	5.78	.51	.91	.92	29.63	20679.10	.22 (.21; .24)
	Women	5.93	.46	.91	.91			
Customer	Men	5.60	.54	.91	.91	30.00	18968.82	.24 (.22; .25)
	Women	5.76	.50	.89	.90			
Accountability	Men	5.41	.52	.86	.86	25.24	20521.34	.19 (.17; .21)
	Women	5.54	.48	.85	.85			
Motivation	Men	5.35	.65	.90	.90	29.18	20252.73	.22 (.20; .24)
	Women	5.53	.61	.90	.90			
Strategy	Men	5.36	.56	.88	.88	17.66	20449.73	.13 (.12; .16)
	Women	5.46	.51	.87	.87			
Innovation	Men	5.50	.54	.88	.88	08.00	19233.19	.06 (.04; .07)
	Women	5.55	.51	.89	.88			

Note. All the effects reported were significant at the .05 level. The results are based on 23,236 male and 14,830 female leaders' scores on the Hogan 360. The following is a brief explanation of important coefficients reported.

- Alpha = Evaluate whether multiple items that measure the same concept produce consistent results. Coefficients above 0.7 are acceptable.
- Omega = Is another measure of scale reliability that is often considered more robust than Alpha. It indicates how well a set of items measures a single underlying concept, and it is beneficial when items have varying levels of strength in measuring the concept.
- t-value = Tells you how strong and clear a result is compared to the natural variation in the data.
- Yuen's d = Measures the size of the difference between two groups, but it ignores extreme values that might impact the results. It gives a more accurate idea of the groups' differences, even when the data do not follow usual statistical assumptions. Values of 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes.

Women obtained statistically significantly higher scores on all the Hogan 360 sub-themes. However, none of the mean group differences were large enough (d > 0.30) to suggest that these differences would be practically observable. Consequently, the evidence suggests that using a single global benchmark for both men and women is unlikely to increase the risk of adverse impact for either gender.

Conclusion

Measurement invariance analyses confirmed that the factor structure of the Hogan 360 is consistent across gender, with equivalent latent factor loadings and satisfactory model fit for both men and women. This indicates that the instrument measures the same constructs across gender groups. Sequential analyses further showed that while women received slightly higher average scores across all subthemes, these differences were minor in magnitude and lacked practical significance (all $d < .30$). These findings align with prior meta-analytic research (e.g., Roth et al., 2012; Mackey et al., 2017), which has reported modest gender differences in performance ratings. Overall, the results provide no evidence of systemic gender bias in Hogan 360 ratings and support the tool's use across gender groups without the need for scoring adjustments or differentiated interpretive frameworks.

References

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

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

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

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. (2024). *A collection of robust statistical methods*. Retrieved from https://cran.r-project.org/web/packages/WRS2/index.html

Mackey, J. D., Roth, P. L., Van Iddekinge, C. H., & McFarland, L. A. (2019). A meta-analysis of gender proportionality effects on job performance. *Group & Organization Management, 44*(3), 578–610. https://doi.org/10.1177/1059601117730519

Nowack, K. M., & Mashihi, S. (2012). Evidence-based answers to 15 questions about leveraging 360-degree feedback. *Consulting Psychology Journal: Practice and Research, 64*(3), 157–182. https://doi.org/10.1037/a0030011

Peter Berry Consultancy (2022). *Hogan 360 technical manual and administration guide*. Peter Berry Consultancy.

R Core Team. (2024). *R: A language and environment for statistical computing.* Retrieved from https:// www.R-project.org/

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. (2024). *Latent variable analysis*. Retrieved from https://cran.r-project.org/web/packages/lavaan/index.html

Roth, P. L., Purvis, K. L., & Bobko, P. (2012). A meta-analysis of gender group differences for measures of job performance in field studies. *Journal of Management, 38*(2), 719–739. https://doi.org/10.1177/0149206310374774

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

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.

Smither, J. W., London, M., & Reilly, R. R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology, 58*(1), 33–66. https://doi.org/10.1111/j.1744-6570.2005.514_1.x

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

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

To cite:

Van Lill, X., & Cruickshank, L. (2026). Technical Brief: Exploring Gender Differences on Hogan 360. Peter Berry Consultancy.

Author Note:

Correspondence regarding this document should be addressed to Peter Berry Consultancy at research@peterberry.com.au

About Peter Berry Consultancy (PBC)

Peter Berry Consultancy (PBC) is a global leadership advisory firm grounded in the science of personality and organisational psychology. We combine rigorous psychometrics with expert consulting to deliver deeper insight into human behaviour at work.

Through evidence-based diagnostics, research, consulting and coaching, we help organisations make better decisions about talent; strengthening leadership, improving team performance, succession planning and enabling sustainable organisational success.

PBC is the developer of proprietary diagnostic tools, including the Hogan 360 Suite, and is an authorised distributor of Hogan Assessments.

Sydney
T: +61 2 8918 0888
Level 8, 201 Miller Street
North Sydney, NSW 2060

Melbourne
T: +61 3 8629 5100
Suite 303, 430 Little Collins
Street, Melbourne, VIC 3000

Auckland
T: +64 9 941 9790
11 Britomart Place, Auckland
CBD, Auckland 1010

Dublin
T: +353 1 578 3607
Suite 301, 53 Merrion Square
South Dublin 2, D02 PR63

www.peterberryconsultancy.com

No part of this work may be copied, reproduced or transferred to any other form or expression without the expressed written consent of Peter Berry Consultancy Pty Ltd.