

# Introduction to Assessment Special Section on Validity Scales in Personality Testing

Robert E. Erard

Received: 13 November 2012 / Accepted: 13 November 2012 / Published online: 5 December 2012  
© Springer Science+Business Media New York 2012

Robert McGrath and colleagues (McGrath, Mitchell, Kim, & Hough, 2010) squarely took aim at a sacred cow in personality assessment when they published a highly provocative meta-analysis in *Psychological Bulletin* that cast doubt on the “validity” of validity scales. Using strict selection criteria, which dramatically winnowed down the number of possibly relevant studies from over 4,000 to 40, they found surprisingly scant evidence supporting the utility of response bias indicators. They concluded that despite close to a century of research devoted to response bias, “the case remains open whether bias indicators are of sufficient utility to justify their use in applied settings to detect misrepresentation” (p. 466). In this and a subsequent article (McGrath, Kim, & Hough, 2011), they issued a challenge for new research that places response bias indicators on a more solid footing.

Alarming as these findings may have sounded to psychologists who routinely rely on validity scales in their daily forensic practice, no one called for an immediate moratorium on their use in the courtroom. Rohling et al. (2011) promptly published a critical response focusing on alleged inadequacies in the methodology of McGrath et al. (2010) and the soundness of their data analysis, particularly with respect to neuropsychological assessment. They argued that McGrath et al. had overlooked at least five studies showing that response bias indicators moderated predictive validity and had made inappropriately sweeping conclusions by treating positive and negative response bias indicators as though evidence concerning the former was relevant to the latter. It is also important to note that the final sample of McGrath et al. (2010) included only one forensic case (i.e.,

Edens & Ruiz, 2006)<sup>1</sup> and only two that directly assessed whether scales are designed to reflect exaggeration of emotional distress operated effectively, thus severely limiting the extent to which the specific findings McGrath et al. could be reliably generalized to forensic cases involving psychological injury.<sup>2</sup>

But the warnings of McGrath et al. about limitations in the research on response bias measures cannot safely be ignored by forensic psychologists (see McGrath et al., 2011). A responsible approach to the challenges they present requires additional reflection and research on whether forensic psychologists have an adequate scientific basis for claiming that the use of validity scale scores provides meaningful evidence about distorted self-presentations concerning psychological symptoms and personal problems on personality tests. This PIL Special Section on Validity Scales in Personality Testing is intended to take a closer look at whether the use of validity scales (particularly those indicating *negative* impression management in personal injury and disability cases) is indeed defensible.

The Special Section opens with Leslie Morey’s detailed examination of the methodology and statistical reasoning behind the arguments in the meta-analysis of McGrath et al. (2010). Morey demonstrates that not only the highly restrictive criteria used by McGrath et al. for selecting studies to include but also the narrowness of the way in which McGrath et al. defined adequate statistical evidence of effective functioning for bias indicators made it extremely improbable that they would find studies that would satisfy them that response bias indicators ever worked as advertised. Morey argues that

---

R. E. Erard (✉)  
Psychological Institutes of Michigan, P.C.,  
7457 Franklin Road, Suite 210,  
Bloomfield Township, MI 48301, USA  
e-mail: rerard2000@ameritech.net

<sup>1</sup> Morey (this issue) reanalyzes the data from this article to show that it actually supports the forensic use of validity scales.

<sup>2</sup> Of course, their failure to find more studies demonstrating the effectiveness of response bias indicators relevant to emotional distress and forensic contexts remains troubling, at least to the extent that one credits the appropriateness of their exclusion criteria.

the methodology of McGrath et al. was, in effect, biased against positive findings. Using several perspicuous examples, he shows how they consequently overlooked substantial positive evidence for validity scales. Focusing on the instrument he knows best, his own personality assessment inventory (Morey, 2007), Morey reviews evidence that Personality Assessment Inventory (PAI) response bias indicators serve both as main effects and as moderators of other substantive predictors.

Next, Wiggins, Wygant, Hoelzle, and Gervais take up the challenge of McGrath et al. (2010, 2011) to researchers to provide new empirical evidence for the effectiveness of validity scales in real-world settings. They conducted an original archival study with a very large sample of actual disability claimants. Their results show that an overreporting group of litigants who scored above the recommended cut scores on multiple validity scales in the MMPI-2-RF Manual (Ben-Porath & Tellegen, 2008) also scored substantially higher than a normal reporting group on all of the restructured clinical (RC) scales. This group accounted for approximately 25 % of the sample, suggesting that symptom exaggeration may be more both common and more detectable than supposed by McGrath et al. (2010). More directly relevant to the McGrath et al. (2011) challenge, Wiggins et al. found that the overreporting group's RC scale scores also showed significantly attenuated validity on many logically relevant external self-report measures.

The last two articles consider the practical utility of negative response bias scales in particular psychological injury contexts. Hoelzle, Nelson, and Arbisi discuss the origins, research base, and utility of a wide range of MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and MMPI-2-RF validity scales, including non-responsiveness, defensive or socially desirable responding, and overreporting of psychological, cognitive, and physical symptoms. They show how particular configurations of these scales can be meaningfully applied in particular litigation settings and integrated with other contextual features to facilitate confident judgments about how personal injury and disability plaintiffs have approached the testing. They also provide illuminating case examples applying these general principles. Finally, they recommend standards for expert testimony concerning motivated distortion in personal injury and disability settings and demonstrate how the use of validity scales can be justified in terms of applicable legal evidentiary standards.

Completing the Special Section, Thomas, Hopwood, Orlando, Weathers, and McDevitt-Murphy<sup>3</sup> offer a similar analysis of PAI validity scales. They review a growing body of empirical evidence concerning the functioning of these scales with plaintiffs claiming to suffer from PTSD. Using a

simulation design, they conducted the largest study to date on the ability of PAI validity scales to distinguish PTSD feigners from normal.<sup>4</sup> They demonstrate that, consistent with several previous studies, all of the PAI validity scales consistently distinguished between two groups, and they found that the validity scales are not very sensitive to coaching. They also report on the first independent test of the Negative Distortion Scale, a new PAI response bias measure that shows particular promise for detection of malingering in PTSD cases. Finally, they offer psychometric data on the relative effectiveness of various validity scales, including guidance on cut scores.

Although more research is clearly needed, showing how, in what circumstances, and to what degree response bias or symptom exaggeration is effectively detected in the forensic use of clinical personality self-report instruments and how standard clinical interpretations should be modified when validity scales are elevated, the articles in this Special Series offer encouraging news. There is reason to believe that McGrath et al. have been excessively conservative in their estimates of the frequency of negative response bias on personality tests and excessively pessimistic about the effectiveness of psychometric methods for measuring it. It is hoped that the articles in this Special Section will stimulate further research that fortifies our knowledge of how to measure and account for deceptive self-presentation in forensic personality assessment.

## References

- Ben-Porath, Y. S., & Tellegen, A. (2008). *Minnesota multiphasic personality inventory-2 restructured form; manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota multiphasic personality inventory-2 (MMPI-2). Manual for administration and scoring*. University of Minnesota Press: Minneapolis.
- Edens, J. F., & Ruiz, M. A. (2006). On the validity of validity scales: The importance of defensive responding in the prediction of institutional misconduct. *Psychological Assessment, 18*, 220–224.
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin, 136*, 450–470.
- McGrath, R. E., Kim, B. H., & Hough, L. (2011). Our main conclusion stands: Reply to Rohling et al. *Psychological Bulletin, 137*, 713–715.
- Morey, L. C. (2007). *Personality assessment inventory professional manual* (2nd ed.). Odessa: Psychological Assessment Resources.
- Rohling, M. L., Larrabee, G. J., Greiffenstein, M. F., Ben-Porath, Y. S., Lees-Haley, P., Green, P., & Greve, K. W. (2011). A misleading review of response bias: Response to McGrath, Mitchell, Kim, & Hough (2010). *Psychological Bulletin, 137*, 708–712.

<sup>4</sup> McGrath et al. (2010) doubted the generalizability of such studies to real-world contexts, but Thomas et al. show that the results of their simulation study are comparable to parallel “real world” studies.

<sup>3</sup> Edited by PIL Assessment Section Co-Head, David Berry