

REPLY

Our Main Conclusion Stands: Reply to Rohling et al. (2011)

Robert E. McGrath
Fairleigh Dickinson University

Brian H. Kim
Occidental College

Leaetta Hough
The Dunnette Group, Ltd

In their comment, M. L. Rohling et al. (2011) accused us of offering a “misleading” review of response bias. In fact, the additional findings they provided on this topic are relevant only to bias assessment in 1 of the domains we discussed, neuropsychological assessment. Furthermore, we contend that, even in that 1 domain, the additional findings they described do not merit revision of our conclusion that the data are insufficient for evaluating the status of bias indicators. We remain hopeful that our review will spur researchers to publish additional tests of the validity of bias indicators in real-world settings and reduce the reliance on analogue studies as an evidence base for their use.

Keywords: response bias, neuropsychology, moderation, malingering

Important decisions affecting people’s lives—decisions with significant personal, financial, and legal implications—are regularly being made based on people’s performance on response bias indicators. We believe that instruments used by psychologists for such purposes must meet stringent standards, and the purpose of our review (McGrath, Mitchell, Kim, & Hough, 2010) was to evaluate bias indicators on the basis of such standards. Rohling et al. (2011) claimed that our results were “misleading.” Even after reviewing the additional research they provided, however, we believe that our primary conclusion—namely, that “sufficient justification for the use of bias indicators in applied settings remains elusive” (p. 450)—still stands.

Rohling et al. (2011) accused us of overgeneralizing but demonstrated overgeneralization themselves when they stated the following in their abstract: “We assert that the relevant and voluminous literature that has addressed the issues of response bias substantiates validity of these indicators” (p. 708). In the text they were somewhat more circumspect. They instead claimed that our conclusion about the ineffectiveness of response bias indicators of positive impression management in organizational settings is probably correct, although they disparaged this finding (and the field of organizational assessment) by referring to it as a “small sample of the psychological assessment world” (p. 709). In fact, hundreds of thousands of people are tested each year in organizational settings

for purposes of personnel selection, placement, and promotion. Despite their claim of effectiveness for bias indicators in most settings, Rohling et al. provided no additional evidence that our conclusions are inaccurate except in the context of neuropsychological assessment of negative impression management.¹

Evidence on Neuropsychological Assessment

The specific focus of Rohling et al.’s (2011) comment is the true status of the response bias hypothesis in the context of neuropsychological assessment. We appreciate their drawing our attention to additional references relevant to the topic. We indicated in our review that it was extremely difficult for us to identify appropriate studies, and we relied heavily on abstracts as a basis for deciding which studies to examine in detail. At no point did we claim that the evidence we cited was sufficient to conclude response bias indicators were ineffective in the context of neuropsychological assessment, only that the evidence we were able to identify was insufficient for a conclusion either way.

Rohling et al. (2011) indicated that the two studies cited in our review on this topic (Bowden, Shores, & Mathias, 2006; Rohling & Demakis, 2010) were suboptimal on two grounds. First, both studies used the “criterion” variable (in the sense of an objective indicator of level of severity) as one of the regression predictors while the regression criterion was the substantive scale, a reversal that we pointed to in note a of our Table 4. We agree that this is a reversal of the correct approach to testing the response bias hypothesis. Nonetheless, although the moderator term of variables

Robert E. McGrath, School of Psychology, Fairleigh Dickinson University; Brian H. Kim, Department of Psychology, Occidental College; Leaetta Hough, The Dunnette Group, Ltd., St. Paul, Minnesota.

We are grateful to Samuel Green for his discussion of some statistical issues raised in this article.

Correspondence concerning this article should be addressed to Robert E. McGrath, School of Psychology T-WH1-01, Fairleigh Dickinson University, Teaneck, NJ 07666. E-mail: mcgrath@fd.edu

¹ Although they referred to their findings as applying to “the context of forensic and disability evaluation” (p. 709) and later implied that their findings apply to the context of pain evaluation when discussing Libon (2010), Rohling et al. (2011) provided no evidence concerning any aspect of disability or forensic evaluation except neuropsychological impairment.

A and B for predicting variable C will not exactly equal the moderator term when A and C are used to predict B, if one moderator effect is small then the other should be small as well. We assert that the failure to find a significant interaction in these eight tests is still evidence against the response bias hypothesis, albeit indirect evidence.

Second, Rohling et al. (2011) indicated that a moderator effect should not be expected in the case of neuropsychological response bias. However, if their argument is correct, bias indicators should cause a suppression effect. Because none of the studies we or Rohling et al. identified have ever examined the suppression effect in neuropsychological assessment, we contend that our original conclusion of insufficient evidence to justify the use of bias indicators remains correct.² Furthermore, the analyses introduced by Rohling et al. examine differences in the strength of the association between the substantive scale and criterion as moderated by a bias indicator. It seems questionable to criticize moderator effects when they contradict one's position but use them as evidence when they support that position.³

Of the five additional studies, one (Gervais, Ben-Porath, Wygant, & Green, 2008) does not meet conditions for our review. We were interested in studies that used a substantive scale and a criterion that was impervious to response bias. The study by Gervais et al. instead involved two substantive scales, a measure of memory complaints and a verbal memory performance task. In the time we had available for this reply we were able to locate only three of the other four studies Rohling et al. (2011) cited (Green, Rohling, Iverson, & Gervais, 2003; Green, Rohling, Lees-Haley, & Allen, 2001; Greiffenstein & Baker, 2003).⁴ These studies reported eight relevant analyses, of which seven were in the direction supportive of the response bias hypothesis. Combining all the results available to us so far, we found 7 of 16 analyses were supportive of the response bias hypothesis.

Even with these additional studies, we continue to assert that the evidence remains insufficient to draw a firm conclusion. There is still not enough information to justify the computation of aggregate effect size estimates, for example. We also note that four of the five studies introduced by Rohling et al. (2011) involved overlapping research teams. Compared to those for normal personality and work-setting assessment, the number of analyses remains small, and more diverse sources of data are needed to protect against allegiance effects (Blair, Marcus, & Boccaccini, 2008).

Final Points

There were several other misrepresentations in the Rohling et al. (2011) comment that merit clarification. We object to their reference to a "confirmatory bias" in our review, as it implies our goal was to undermine response bias measures. A close reading of our article does not support that implication. We indicated that we consider the detection of response bias an essential activity; we noted promising, albeit insufficient, evidence for the value of variable response inconsistency scales (Tellegen, 1988); and we concluded by suggesting that multiple independent sources of evidence for biased responding may be a more effective approach to detecting bias than single measures (although this topic is also underresearched). However, we stand by our original conclusion, that our review "raises concerns about the validity of bias indica-

tors in those settings where sufficient research exists to draw a conclusion and the justification for their use in those settings where the research is insufficient" (McGrath et al., 2010, p. 465).

Rohling et al. (2011) also claimed that we attempted to "subvert" the scientific review process by thanking those individuals who provided us with materials. Someone reading the list of individuals we thanked would surely have also read the next sentence indicating that those we thanked did not necessarily agree with the conclusions of our study. The note expressed our gratitude for help with a very complex study and nothing more.

For far too long, the justification for using response bias measures has depended too heavily on analogue research. The ease of conducting such studies makes them attractive, but it is an insufficient standard when one considers the potential consequences of false positives. Although Rohling et al. (2011) have contributed to the issue by bringing additional studies to light, the need remains for additional research that justifies the use of bias indicators in real-world applications.

² The term *suppression* is sometimes used in the neuropsychological literature to refer to the proposition that poor effort reduces scores on substantive scales. We want to be clear that we are talking about suppression in the statistical sense described in McGrath et al. (2010).

³ Rohling et al. (2011) used the term *interaction* rather than moderation, so we considered the possibility they meant that moderation effects could exist but interaction effects as a specific method of testing for moderation would be nonsignificant. Because our review was not restricted to interaction effects, we assume that they are discussing moderation effects in general.

⁴ We decided against including the results from the missing article, Green (2007), in our conclusions as they were reported by Rohling et al. (2011), because we disagreed with their characterization of three of the articles we located. As noted, we would not have included Gervais et al. (2008). Rohling et al. (2011) described two analyses from the Greiffenstein and Baker (2003) reference, but there were actually four analyses relevant to the response bias hypothesis, one of which contradicted the hypothesis. Although Green et al. (2003) described four criteria, they provided information about putative fakers only for three.

References

- Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science and Practice, 15*, 346–360. doi:10.1111/j.1468-2850.2008.00147.x
- Bowden, S. C., Shores, E. A., & Mathias, J. L. (2006). Does effort suppress cognition after brain injury? A re-examination of the evidence for the Word Memory Test. *Clinical Neuropsychologist, 20*, 858–872. doi: 10.1080/13854040500246935
- Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *Clinical Neuropsychologist, 22*, 1061–1079. doi:10.1080/13854040701756930
- Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America, 18*, 43–68. doi:10.1016/j.pmr.2006.11.002
- Green, P., Rohling, M. L., Iverson, G. L., & Gervais, R. O. (2003). Relationships between olfactory discrimination and head injury severity. *Brain Injury, 17*, 479–496. doi:10.1080/0269905031000070242
- Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compen-

- sation claimants. *Brain Injury*, 15, 1045–1060. doi:10.1080/02699050110088254
- Greiffenstein, M. F., & Baker, W. J. (2003). Premorbid clues? Preinjury scholastic performance and present neuropsychological functioning in late postconcussion syndrome. *Clinical Neuropsychologist*, 17, 561–573. doi:10.1076/clin.17.4.561.27937
- Libon, D. J. (2010). Neurobiological aspects of complex regional pain syndrome (CRPS): Reply to Victor, Boone, and Kulick (2010). *Journal of the International Neuropsychological Society*, 16, 1153–1154. doi:10.1017/S1355617710001049
- 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. doi:10.1037/a0019216
- Rohling, M. L., & Demakis, G. J. (2010). Bowden, Shores, & Mathias (2006): Failure to replicate or just failure to notice. Does effort still account for more variance in neuropsychological test scores than TBI severity? *Clinical Neuropsychologist*, 24, 119–136. doi:10.1080/13854040903307243
- 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: Comment on McGrath, Mitchell, Kim, and Hough (2010). *Psychological Bulletin*, 137, 708–712.
- Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality*, 56, 621–663. doi:10.1111/j.1467-6494.1988.tb00905.x

Received March 14, 2011

Revision received March 18, 2011

Accepted March 21, 2011 ■