Development of a Response Bias Scale for the MMPI-2

Roger Gervais, Ph.D., Neurobehavioural Associates, Edmonton, Alberta
2005 Annual Meeting of the National Academy of Neuropsychology, Tampa, FL., October 19-22, 2005.

Abstract

Objective: To derive a Response Bias Scale (RBS) for the MMPI-2 using empirical methods.

Method: This archival study examined Word Memory Test (WMT; Green, 2003; Green, Allen, & Astner, 1996; Green & Astner, 1995), Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997), and MMPI-2 raw data from 1212 consecutive non-head injury disability claimants and counseling clients seen in a private practice setting. Logistic and Multiple Regression analyses were used with the MMPI-2 items as independent variables and pass/fail WMT and/or CARB as a grouping criterion to select MMPI-2 items that predicted group membership.

Results: Regression analyses identified 41 MMPI-2 items that classified claimants as pass/fail WMT with 78% overall accuracy in the total sample, and with 82% accuracy in a chronic pain disability subgroup. The 41-item scale (RBS) correlated with the mean of the three WMT effort measures (r = -.50, p < .0005). The scale was highly accurate in identifying SVT failure in the total sample and in subgroups with specificity between 92% and 96%, and positive predictive power ranging from 74% to 92% at a cutoff of 22. Positive predictive power was 100% at a score greater or equal to 27. The RBS demonstrated significant incremental validity above the F scale in predicting WMT and MSVT failure in the primary sample and in a cross-validation sample.

Conclusions: Preliminary analyses of the RBS suggest that it is a potentially useful measure of response bias in the cognitive, emotional, and physical symptom dimensions. Further cross-validation with different clinical samples is in progress.

Materials and Methods

Participants

This study used archival data from 1212 consecutive non-head injury disability claimants and counseling clients referred to the author's private psychology practice. The sample was 51% male, 40.9 (10.6) years of age, and had a mean education level of 12 (2.6) years. WCB (54%) and legal (22%) referrals constituted the majority of the sample. Approximately 39% of cases had a diagnosis of chronic pain, 23% had anxiety-related diagnoses, 17% had orthopedic diagnoses, and 15% presented with depression as the primary diagnosis.

Assessment Methods

All persons in this study were administered a psychological assessment battery consisting of a variety of cognitive tests, the MMPI-2, and self-report symptom questionnaires. Eighty-nine percent of cases completed the Word Memory Test (WMT; Green, 2003; Green, Allen & Astner, 1996; Green & Astner, 1995) and 98% completed the Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997). A total of 59% of the claimants also completed the TOMM (Tombaugh, 1996). Cutoffs for determining failure on the SVTs were set in accordance with the respective test manuals. The WMT registered the highest failure rate at 32%, followed by the CARB at 17% and the TOMM at

13% of the sample, according to standard failure criteria. A total of 39% of the sample failed one or more of the SVTs. A small subset of the sample also completed the Medical Symptom Validity Test (formerly known as the Memory and Concentration Test) (MSVT; Green, 2004). The True/False MMPI-2 item responses for each case were manually entered into the database. Seventy-five percent of cases completed the full 567-item MMPI-2. Dustin Wygant (Kent State University) scored the resulting dataset, rendering the standard validity and clinical scales, as well as a number of other measures and indices. In the initial analyses, we did not exclude any cases on the basis of MMPI-2 profile validity criteria.   Standard MMPI-2 exclusion criteria (CNS $\geq$ 30, TRIN/VRIN $\geq$80) were implemented in the incremental validity analyses, resulting in a final sample size of 775 cases who completed the full MMPI-2 and the WMT (see Gervais, Wygant, & Ben-Porath, 2005).

<u>Statistical Analyses</u>

Logistic and Multiple Regression analyses were used to identify a pool of MMPI-2 items that best classified the sample into pass/fail WMT or CARB groups. Preliminary analyses indicated that the majority of predictor items were in the first 370 items. Subsequent analyses were limited to the 370-item form of the MMPI-2, as we felt this would be of the greatest practical utility to clinical practitioners.

Incremental validity of the resulting 41-item Response Bias Scale (RBS) relative to the MMPI-2 F scale in predicting SVT failure was evaluated by means of linear regression analyses. These analyses were repeated with a cross-validation sample (N = 222).   The final analysis involved calculating the sensitivity, specificity, positive predictive power, and negative predictive power of the scale relative to predicting WMT and/or MSVT failure.

<center>Results</center>

Logistic regression analysis using the total sample (N = 1212) with pass/fail WMT and/or CARB as the dependent variable and the first 370 MMPI-2 items as the predictor variables produced a significant model (chi-square = 770.64, df = 370, p < 0.0005). The model accounted for between 51.0% and 69.2% of the variance in WMT/CARB failure status.   Overall, the model was 87.1% accurate, correctly classifying 90.8% cases that passed WMT and CARB, and 81.3% of WMT/CARB failures.

Excluding all non-significant predictor variables left 78 MMPI-2 items with which we repeated the analysis.   Non-significant variables were excluded from subsequent analyses until we achieved a model in which there were 36 significant MMPI-2 items, producing an overall 75.4% classification accuracy (chi-square = 351.16, df = 36, p < 0.0005).

As a final step in the analysis, we conducted a stepwise multiple regression using all significant MMPI-2 items from the above analyses as the predictor variables, with pass/fail WMT/CARB as the dependent variables. The analysis produced a model with 24 significant predictor variables (adjusted R square = .22; F = 10.9, p < 0.0005).   We repeated the analyses with the same pool of predictor variables, using pass/fail only WMT (adjusted R square = .21; F = 11.2, p < 0.0005), and fail WMT with a modified 70% cutoff (adjusted R square = .17; F = 12.7, p < 0.0005).   We combined the significant MMPI-2 items from the three analyses with the 36 items derived from the logistic analyses. There were five non-overlapping items between the two item sets, leading to a final 41-item scale (RBS), which produced an overall pass/fail WMT classification accuracy rate of 78.3%, and an 82%

classification rate in a chronic pain subgroup. The 41-item scale (RBS) correlated with the mean of the three WMT effort measures (r = -.50, p < .0005).

RBS scores in the sample ranged from 6 – 31 (M = 17.1, SD = 4.4, median = 17). The mean score of cases who passed WMT and CARB was 15.6 (SD = 3.6, median = 16), and 19.9 (SD = 4.1, median = 20) in those who failed WMT and/or CARB.   There were no gender-based differences in the mean RBS scores.

Correlations between the RBS scale and other symptom validity measures in the test battery were moderately strong ranging from .34 for F_P (Arbisi & Ben-Porath, 1995) to .63 for the Meyers Validity Index (MVI; Meyers, 2002)(p < 0.0005). Details of the correlations between RBS and other SVTs and response bias measures in the battery are contained in Table 1.

Incremental Validity of RBS

Butcher, Graham, and Ben-Porath (1995) emphasize that the incremental validity of new scales must be evaluated to determine whether the scale adds significantly to the prediction of the behavior in question.   Following the methodology of Arbisi and Ben-Porath (1995), we conducted linear regression analyses to evaluate the incremental validity of the RBS compared to the F scale, the traditional index of symptom exaggeration or malingering on the MMPI-2 (Rogers, 1997), in discriminating between the pass/fail WMT groups. In the first analysis, pass/fail WMT group membership (dependent variable) was regressed onto the F and RBS scales (independent variables). The F scale T score was entered first, followed by the RBS raw score, resulting in a significant model accounting for 18% of the variance. The ability of RBS to contribute incrementally to the prediction of group membership was determined by the F (change) statistic. In the second analysis the order of entry was reversed, with RBS entering the regression equation first, followed by the F scale in the second block.   The F (change) statistic indicated the incremental contribution of the F scale to contribute to the prediction of pass/fail WMT group membership. The results of these analyses are presented in Table 2.

Examination of Table 2 reveals that RBS added significantly to the F scale (17% of the variance, p < 0.0005) in the prediction of pass/fail WMT group membership. Conversely, while still statistically significant, the F scale contributed minimally (1% of the variance) to the prediction of group membership. This is illustrated further by the beta weights associated with the regression analyses (.48 and -.12, RBS and F, respectively).   In these analyses, therefore, RBS provided the most predictive power.

Recognizing that there are methodological problems associated with validating a scale using the same sample from which it was derived, we repeated the above regression analyses using pass/fail TOMM group membership as the dependent variable.   Whereas the WMT had been used as a criterion variable in analyses to select the RBS items, the TOMM is a completely separate and distinct SVT, which was not used in the development of the scale.   The sample contained 548 consecutive cases that had been administered the booklet form of the TOMM. The regression analysis produced a significant model accounting for 15% of the variance. When F was entered first, RBS accounted for 13% of the variance (p < 0.0005). Reversing the order of entry found that F contributed less than 1% of the variance above RBS in predicting TOMM failure.   Review of the beta weights confirmed that RBS again provided most of the predictive power (Table 3).

Cross-validation of RBS

Cross-validation of the RBS scale was undertaken with a sample of patients from two independent clinical practices (N = 222). The first sample was comprised of a consecutive series of 141 persons referred to the author's practice for psychological assessment. The sample was 57% male, 41 (11.0) years of age, and had 11.9 (2.6) years of education. A total of 83% of the persons were involved in some type of disability claim or litigation (63% WCB). The majority of cases had anxiety-related diagnoses (41%) with chronic pain and depression as primary diagnoses in 28% and 27% of the sample, respectively. Nearly all cases (n = 126) had completed the full 567 MMPI-2, with the remaining 15 completing the 370-item form.

The second sample consisted of 81 consecutive patients referred for neuropsychological assessment to the practice of Dr. Paul Green in Edmonton, Alberta. The Green sample was 58% male, 43.4 (10.8) years of age, with a mean of 12.5 (2.8) years of education. The majority of cases were involved in some form of disability claim. The largest diagnostic class was composed of persons with head injuries (mild = 21%, moderate to severe = 20%). Miscellaneous diagnoses (21%) and depression (16%) were the primary diagnoses in the remainder of the sample. The persons in this sample completed only the abbreviated 370-item form of the MMPI-2.

We implemented the above-noted exclusion criteria (CNS $\geq$ 30, VRIN/TRIN $\geq$ 80) in cases who completed the full 567 MMPI-2 form. For the Green sample and other cases who completed only the 370 items, we used a prorated CNS $\geq$ 20. Seven cases failed the exclusion criteria resulting in final cross-validation sample size of 215. Considering that the F scale is derived from the first 370 items of the MMPI-2, the modified exclusion criteria allowed us to use the Green sample and maximize the statistical power for the analyses.

Table 4 presents the results of incremental validity analyses repeated with the cross-validation sample. When F was entered first into the regression equation, RBS accounted for an additional 9% of the variance (p < 0.0005) in predicting pass/fail WMT group membership. Entering RBS first accounted for 14% of the variance, with F explaining less than 1% of the additional variance, a non-significant increment. Similar findings were obtained with F and RBS regressed onto pass/fail MSVT, CARB, and TOMM group membership as described in Tables 5-7. Review of the beta weights for these analyses supported our conclusion that RBS had the greatest power in predicting group membership.

Positive Predictive Power

A receiver operating characteristic (ROC) curve analysis was used to evaluate possible RBS cutoffs in most effectively predicting failure on WMT, CARB, or TOMM in the primary sample. The area under the curve (AUC) of .78 (CI = .75-.81) was reasonably good. A cutoff score of 22 on the RBS was associated with an acceptable false positive rate of .05 and a true positive rate of .35. Table 8 presents prevalence of WMT or MSVT failure, sensitivity, specificity, positive and negative predictive power and hit rate for the RBS in the cross-validation sample at a cutoff score of 22. Values for the total sample as well as diagnostic subgroups are also provided. Positive predictive power ranged from .71 to .89 in the anxiety and depression subgroups, respectively. PPP in the chronic pain subgroup was strong at .86.

External Validity

Finally, any measure of response bias should predict clinically significant effects on other measured behaviors. It is well established that biased responding or poor effort on SVTs is associated with under performance on objective cognitive tests, and general symptom exaggeration or over reporting on self-report questionnaires (Green, Rohling, Lees-Haley, & Allen, 2001). If externally valid as a measure of response bias/effort, increasing RBS scores should also be associated with significant changes in the target test results. This assumption was tested by a series of t tests contrasting RBS raw scores within the 1st and 4th quartiles (0-14, 21+) with performance on various cognitive test scores and self-report ratings. Significant differences ($p < 0.0005$) were found between the two RBS score ranges for all test measures. Effect sizes (Cohen's d) were medium to very large (-.61 for delayed story recall, to 2.10 for the mean Memory Complaints Inventory score) (MCI; Green, Gervais, & Merten, 2005; Green, 2004; Green & Allen, 1997) as presented in Table 9. This supports the interpretation that elevated scores on the RBS are associated with response bias or incomplete effort.

Discussion

The present study undertook to develop a response bias scale for the MMPI-2 using regression analyses to identify items that predicted failure on the WMT and CARB. The resulting 41-item scale demonstrated significant incremental validity above the MMPI-2 F scale in predicting failure on the WMT, MSVT, CARB, and TOMM in both the original sample and in a cross-validation sample. Sensitivity, Specificity, PPP, and NPP in predicting SVT failure were generally greater for the RBS in the cross-validation sample than the MMPI-2 F, F_B, F_P scales and other indices of biased responding in the primary sample (see Gervais, Wygant, Ben-Porath, 2005). Increasing RBS scores were associated with poorer performance on objective cognitive tests and greater symptom reports. These findings suggest that the RBS is a potentially useful index of response bias as measured by the WMT and other measures of predominantly cognitive response bias. Possible application of the scale include situations in which formal measures of cognitive response bias were not administered, or where relatively insensitive SVTs were employed or coaching is suspected. This is particularly relevant given the recent NAN position paper on the need for symptom validity assessment as a medical necessity in neuropsychological assessments (Bush, Ruff, Tröster, et al., 2005).
The present study was based on a sample of predominantly non-head injury disability claimants. The cross-validation sample contained a significant number of persons referred for neuropsychological assessment of head injury or other neurological conditions. Further research is needed with larger samples of persons with different diagnoses, with and without disability incentives.

References

Allen, L., Conder, R. L., Green, P., & Cox, D. R. (1997). CARB' 97 Manual for the Computerized Assessment of Response Bias. CogniSyst, Inc., Durham, NC.

Arbisi, P. & Ben-Porath, Y. (1995). An MMPI-2 infrequency response scale for use with psychopathological populations: the Infrequency Psychopathology Scale, F(p). Psychological Assessment, 7, 424-431.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, D. R., Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity. NAN

Policy & Planning Committee. Archives of Clinical Neuropsychology, 20, 419-426.

Butcher, J. N., Graham, J. R., & Ben-Porath, Y. (1995). Methodological problems and issues in MMPI, MMPI-2, and MMPI-A Research. Psychological Assessment, 7, 320-329.

Gervais, R. O., Wygant, D. B., & Ben-Porath, Y. S. (2005). Word Memory Test (WMT) performance and MMPI-2 validity scales and indices. Presented at Annual conference of National Academy of Neuropsychology, Tampa, Florida, October 2005. Archives of Clinical Neuropsychology, 20 (7), 891-92.

Gough, H. G. (1950). The F minus K dissimulation index for the MMPI. Journal of Consulting Psychology, 14, 408-413.

Gough, H. G. (1957). California psychological inventory manual. Palo Alto, CA: Consulting Psychologists Press.

Green, P. (2004). Green's Medical Symptom Validity Test (MSVT): User's Manual. Edmonton: Green's Publishing.

Green, P. (2004). The MemoryComplaints Inventory (MCI) for Windows. Edmonton: Green's Publishing.

Green, P. (2003). Green's Word Memory Test for Windows: User's manual. Edmonton: Green's Publishing.

Green, P. & Allen, L. (1997). Memory Complaints Inventory. Durham, NC: CogniSyst, Inc.

Green, P. & Astner, K. (1995). Manual for the Oral Word Memory Test. Durham, NC: CogniSyst, Inc.

Green, P., Allen, L., & Astner, K. (1996). The Word Memory Test: A User's Guide to the Oral and Computer-Administered Forms, US Version 1.1. Durham, NC: CogniSyst, Inc.

Green, P., Gervais, R., & Merten, T. (2005). Das Memory Complaints Inventory (MCI): Gedächtnisstörungen, Berschwerdenschilderung und Leistungsmotivation [The Memory Complaints Inventory (MCI): Memory impairment, symptom presentation, and test effort]. Neurologie & Rehabilitation, 11 (3), 139-144.

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 compensation claimants. Brain Injury, 15(12), 1045-1060.

Lees-Haley P. R., English L.T., & Glenn W.J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. Psychological Reports, 68, 203-210.

Meyers, J. E., Millis, S. R., & Volkert, K. (2002). A validity index for the MMPI-2. Archives of Clinical Neuropsychology, 17, 157-169.

Rogers, R. (1997). Clinical assessment of malingering and deception, 2nd edition. New York: The Guilford Press.

Steffan, J., Clopton. J., & Morgan, R. (2003). An MMPI-2 scale to detect malingered depression (Md Scale). Assessment, 10, 383-392.

Tombaugh, T. N. (1996). Test of Memory Malingering. Toronto, ON: MultiHealth Systems.

Wygant, D. B. (2004). Development and initial validation of a scale to detect infrequent somatic complaints. Unpublished master's thesis, Kent State University, Kent, Ohio.

Appendix

Response Bias Scale (RBS) 41-item scale composition, © Roger O. Gervais, 2005

Count one point if scored in the keyed direction:

True: 31, 40, 94, 97,147, 153, 156,168, 175, 192, 229, 253, 269, 282, 283, 303, 308, 309, 310, 319, 351

False: 25, 35, 42, 43, 46, 78, 89, 105, 109, 137, 169, 185, 186, 260, 264, 315, 330, 340, 354, 362

Table 1. Correlations between RBS and MMPI-2 validity scales and indices (N = 775)

F	F_B	F_P	F_S	F-K	Ds-r	FBS	MVI	Md
.53	.56	.34	.56	.46	.48	.60	.63	.59

Note. F_P (Arbisi & Ben-Porath, 1995), F_S (Wygant, 2004), F-K (Gough, 1950), Ds-r (Gough, 1957); FBS (Lees-Haley, English, & Glenn, 1991), MVI (Meyers, Millis, & Volkert, 2002), Md (Steffan, Clopton, & Morgan, 2003).

Table 2. Incremental validity: Discriminating between pass and fail WMT groups in main sample (N = 775)

Variable	R	R^2_{change}	F_{change}	P_{change}	Final Beta
			Entering F first		
F	.14	.02			-.12
RBS41	.43	.17	156.31	<0.0005	.48
			Entering RBS first		
RBS41	.42	.18			.48
F	.43	.01	9.04	.003	-.12

Table 3. Incremental validity: Discriminating between pass and fail TOMM groups in main sample (N = 548)

Variable	R	R^2_{change}	F_{change}	P_{change}	Final Beta
			Entering F first		
F	.15	.02			.09
RBS41	.39	.13	82.33	<0.0005	.43
			Entering RBS first		
RBS41	.14	.14			.43
F	.15	.005	3.40	.07	.09

Table 4. Incremental validity: Discriminating between pass and fail WMT groups in cross-validation sample (N = 215)

Variable	R	R^2_{change}	F change	P change	Final Beta
			Entering F first		
F	.24	.06			.05
RBS41	.38	.09	20.94	<0.0005	.35
			Entering RBS first		
RBS41	.38	.14			.35
F	.38	.002	.49	.48	.05

Table 5. Incremental validity: Discriminating between pass and fail MSVT groups in cross-validation sample (N = 215)

Variable	R	R^2_{change}	F change	P change	Final Beta
			Entering F first		
F	.05	.05			-.03
RBS41	.20	.15	36.63	<0.0005	.46
			Entering RBS first		
RBS41	.45	.20			.46
F	.45	.001	.16	.69	-.03

Table 6. Incremental validity: Discriminating between pass and fail CARB groups in cross-validation sample (N = 215)

Variable	R	R^2_{change}	F change	P change	Final Beta
			Entering F first		
F	.21	.05			-.02
RBS41	.43	.14	31.51	<0.0005	.44
			Entering RBS first		
RBS41	.43	.18			.44
F	.43	.000	.06	.81	-.02

Table 7. Incremental validity: Discriminating between pass and fail TOMM groups in cross-validation sample (N = 215)

Variable	R	R^2_{change}	F change	P change	Final Beta
			Entering F first		
F	.25	.06			.04
RBS41	.46	.15	22.06	<0.0005	.45
			Entering RBS first		
RBS41	.46	.21			.45
F	.46	.001	.14	.71	.04

Table 8. RBS classification rates at cutting score of 22 distinguishing pass fail WMT or MSVT in various diagnostic classes in cross-validation sample (N = 215)

	Prevalence	Sensitivity	Specificity	PPP	NPP	Hit Rate
Total sample	.39	.34	.92	.74	.69	.69
Chronic Pain	.44	.38	.95	.86	.66	.69
Depression	.51	.36	.95	.89	.59	.65
Anxiety	.34	.29	.94	.71	.72	.72

Table 9. Cognitive and self-report scores by RBS 1st and 4th quartile levels

	RBS Level						
	1st Quartile 0-14 (n =250)		4th Quartile 21+ (n =257)		t test		Effect Size Cohen's d
	M	SD	M	SD	t	p	d
Full Scale IQ	101.2	12.3	92.1	14.3	5.5	< .0005	-.68
CVLT SDFR	11.1	3.1	8.5	3.5	6.8	< .0005	-.78
SRT Immed.	46.5	9.1	37.8	11.8	7.0	< .0005	-.83
SRT Delay	33.1	13.4	24.8	13.8	3.7	< .0005	-.61
EPT	12.0	4.6	16.1	6.7	4.9	< .0005	.71
Smell ID	5.9	1.2	4.6	2.1	4.5	< .0005	-.64
BDI	17.1	9.3	34.9	10.9	17.1	< .0005	1.73
BAI	15.4	10.1	31.7	12.6	11.9	< .0005	1.41
MPI Pain	42.2	11.5	52.6	8.9	10.8	< .0005	1.03
MCI Mean	15.1	11.0	48.0	18.6	20.4	< .0005	2.10
MCI-IRM	10.3	11.6	28.7	19.0	11.1	< .0005	1.14
MCI-ACB	6.8	10.0	39.6	23.3	17.5	< .0005	1.76
MCI-AAB	4.1	10.2	20.3	19.0	10.1	< .0005	1.03
DAPS-PB	44.9	11.7	37.3	7.3	4.1	< .0005	-.78
DAPS-NB	53.4	12.1	88.0	39.6	6.2	< .0005	1.18

Note. SRT (Story Recall Test; Immediate and Delayed), EPT (Emotional Perception Test), Smell Identification (Alberta Smell Test), BDI (Beck Depression Inventory), BAI (Beck Anxiety Inventory), MPI (Multidimensional Pain Inventory), MCI (Memory Complaints Inventory; Impairment of Remote Memory, Amnesia for Complex Behavior, Amnesia for Antisocial Behavior), DAPS (Detailed Assessment of Posttraumatic Stress; Positive Bias, Negative Bias).

Contact information:

Roger Gervais, Ph.D.
Suite 201, 17107-107 Avenue, Edmonton Alberta, Canada
T5S 1G3
E-mail: rgervais@shaw.ca