

Use of MMPI-2 to predict cognitive effort: A hierarchically optimal classification tree analysis

COLETTE M. SMART,¹ NATHANIEL W. NELSON,^{2,3} JERRY J. SWEET,^{4,5} FRED B. BRYANT,⁶
DAVID T.R. BERRY,⁷ ROBERT P. GRANACHER,⁸ AND ROBERT L. HEILBRONNER^{5,9}

¹Department of Cognitive Rehabilitation, JFK-Johnson Rehabilitation Institute, Edison, New Jersey

²Psychology Service, Minneapolis VA Medical Center, Minneapolis, Minnesota

³Department of Psychiatry, University of Minnesota, Minneapolis, Minnesota

⁴Department of Psychiatry & Behavioral Sciences, Evanston Northwestern Healthcare, Evanston, Illinois

⁵Feinberg School of Medicine, Northwestern University, Evanston, Illinois

⁶Department of Psychology, Loyola University Chicago, Chicago, Illinois

⁷Department of Psychology, University of Kentucky, Lexington, Kentucky

⁸Lexington Forensic Institute, Lexington, Kentucky

⁹Chicago Neuropsychology Group, Chicago, Illinois

(RECEIVED December 5, 2007; FINAL REVISION May 12, 2008; ACCEPTED May 23, 2008)

Abstract

Neuropsychologists routinely rely on response validity measures to evaluate the authenticity of test performances. However, the relationship between cognitive and psychological response validity measures is not clearly understood. It remains to be seen whether psychological test results can predict the outcome of response validity testing in clinical and civil forensic samples. The present analysis applied a unique statistical approach, classification tree methodology (Optimal Data Analysis: ODA), in a sample of 307 individuals who had completed the MMPI-2 and a variety of cognitive effort measures. One hundred ninety-eight participants were evaluated in a secondary gain context, and 109 had no identifiable secondary gain. Through recurrent dichotomous discriminations, ODA provided optimized linear decision trees to classify either sufficient effort (SE) or insufficient effort (IE) according to various MMPI-2 scale cutoffs. After “pruning” of an initial, complex classification tree, the Response Bias Scale (*RBS*) took precedence in classifying cognitive effort. After removing *RBS* from the model, *Hy* took precedence in classifying IE. The present findings provide MMPI-2 scores that may be associated with SE and IE among civil litigants and claimants, in addition to illustrating the complexity with which MMPI-2 scores and effort test results are associated in the litigation context. (*JINS*, 2008, 14, 842–852.)

Keywords: Malingering, Neuropsychological assessment, Optimal discriminant analysis, Personality assessment, Response validity, Response Bias Scale (*RBS*)

INTRODUCTION

Psychological and cognitive response validity measures are often administered concurrently in secondary gain (SG) contexts to provide greater understanding with regard to the veracity of individual neuropsychological performances. Regarding psychological response validity measures, the Minnesota Multiphasic Personality Inventory (MMPI-2; Butcher et al., 1989) has been the most widely examined

instrument in this area of research, and depending upon the SG setting, MMPI-2 profiles may represent “under-reporting” or “over-reporting” of symptoms. For example, clinicians that administer the MMPI-2 as part of a hiring process (Pope et al., 2000) or in the context of custody litigation (Posthuma & Harper, 1998) may reasonably expect respondents to have characteristic underreporting validity and clinical profiles. Conversely, other studies have examined whether select MMPI-2 validity scales (e.g., the *F*-family: *F*, *Fb*, *Fp*) and clinical scales (e.g., *Hs*, *D*, *Hy*, *Pt*, *Sc*) may be differentially sensitive to over-reporting of symptoms in secondary gain (SG) contexts, such as personal injury litigation. Whereas the *F*-scale and *Fp* (Arbisi &

Correspondence and reprint requests to: Jerry J. Sweet, Neuropsychology Service, Department of Psychiatry and Behavioral Sciences, Evanston Northwestern Healthcare Medical Group, 909 Davis Street, Suite 160, Evanston, IL 60201. E-mail: j-sweet@northwestern.edu

Ben-Porath, 1995) have been found to be quite sensitive in identifying “rare-symptoms” endorsed by over-reporting and comparison groups (Rogers et al., 2003), other researchers have found *F* and related scales to be less effective than postrelease scales, such as the Fake Bad Scale (*FBS*; Lees-Haley et al., 1991) in identifying response bias among neuropsychological civil litigants (e.g., Larrabee, 2003). Moreover, examination of clinical scale profiles in addition to validity scale profiles is important in SG contexts considering that some SG groups, such as litigants, may demonstrate clinical scale elevations (e.g., *Hs*, *Hy*) in the absence of significantly elevated validity scale elevations (Lanyon & Almer, 2002).

As regards cognitive response validity assessment, a substantial literature has documented the sensitivity of various effort measures in SG contexts (see Bianchini et al., 2001), and forced-choice effort measures are among the most widely administered measures in neuropsychological practice. Less-than-chance performance on forced-choice measures has been suggested as strongly increasing one’s confidence in arriving at a diagnosis of malingering (Slick et al., 1999), though cut-offs above the chance-level may also implicate insufficient effort (IE). For instance, the “90% rule” (i.e., raw score cutoffs that are less than 90% correct on forced-choice measures) is a commonly suggested “rule of thumb” and raises the possibility of IE (e.g., Grote et al., 2000). A previous meta-analysis (Vickery et al., 2001) found the Digit Memory Test (DMT; Hiscock & Hiscock, 1989) to be most effective in identifying IE relative to other effort measures examined, including non-forced-choice measures. Similar forced-choice measures, such as the Victoria Symptom Validity Test (VSVT; Slick et al., 1995), the Test of Memory Malingering (TOMM; Tombaugh, 1996), Multi-Digit Memory Test (MDMT; Niccolls & Bolter, 1991), Word Memory Test (WMT; Green, 2003), and the Letter Memory Test (LMT; Inman et al., 1998) have also demonstrated respective utility in civil and simulating samples.

While response validity research has demonstrated relative progress in evaluation of psychological and cognitive response validity measures in their own right, relatively few studies have examined whether MMPI-2 validity scales can be expected to improve *clinical decision-making* with regard to cognitive effort. The literature in this area has been somewhat equivocal to date, with some studies suggesting the potential that some psychological response validity scales may moderate cognitive effort (e.g., Gervais, 2005; Nelson et al., 2007a), and others suggesting relative independence of psychological *versus* cognitive response validity (e.g., Greiffenstein et al., 1995). In the latter exploratory factor analysis (EFA), the authors found that MMPI-2 and cognitive response validity measures loaded on discrete factors, suggesting the possibility of relatively minimal overlap among psychological and cognitive response validity constructs. A more recent EFA (Nelson et al., 2007a) suggests that the relationship between psychological and cognitive response validity variables is complex: while cognitive effort loaded independently from factors associated with

over-reporting of psychological symptoms in general, validity scales whose content reflected over-reporting of somatic/neurotic symptoms (e.g., *FBS*) had a greater relationship with cognitive effort than over-reporting of psychotic symptoms (e.g., *F*, *Fp*, *F-K*).

One recent postrelease scale, the Response Bias Scale (*RBS*; Gervais, 2005), was developed with the explicit intention of identifying MMPI-2 items that might be particularly relevant to cognitive effort in civil forensic groups. Specifically, non-head-injury claimant effort performances were obtained on the Word Memory Test (Green, 2003), and MMPI-2 test items showing particular discrimination of sufficient effort (SE) and IE groups contributed to *RBS* development. WMT effort performances showed a relative decrease in performance with increasing *RBS* magnitudes. In an independent study, *RBS* showed preliminary merit by demonstrating a moderate effect size ($d = .65$) between SG and non-SG clinical groups (Nelson et al., 2007b). Although cognitive effort was not examined in the latter study, *RBS* was among the MMPI-2 validity scales to load on the “over-reporting of neurotic symptoms” factor in the Nelson and others’ (2007a) EFA. This factor demonstrated a greater correlation with cognitive effort than the “over-reporting of psychotic/rarely endorsed symptoms” factor, which provides preliminary support of the notion that *RBS* and other validity scales whose content reflects “somatic” symptoms (e.g., *FBS*) might have a unique relationship with cognitive effort.

However, response validity research is most beneficial when it impacts the clinician’s everyday practice. EFA, while documenting a possible association among certain psychological and cognitive response validity constructs, is not necessarily the most clinically relevant approach to response validity research, and a variety of other methodological strategies in the context of response validity research are of potentially greater clinical relevance. Provision of base rate MMPI-2 data in sizeable SG groups (e.g., Lees-Haley, 1997; Mittenberg et al., 2002) allows for an understanding of general response validity trends in SG groups, even if it does not provide a thorough dissemination of how or why these trends may be present. Another strategy includes “simulation” research, whereby certain groups are given coached instructions to over-report symptoms on the MMPI-2 or effort measures and comparison groups are given standard instructions (e.g., Bagby et al., 2000; Dearth et al., 2005; Rogers et al., 1995). Results of the coached and uncoached groups are then contrasted according to clinically relevant cut-scores, and the clinician is provided with known classification accuracy rates (e.g., specificity, sensitivity, positive and negative predictive validity at varying base rates of malingering). A “known groups” methodology is thought to better account for the “real-world” nature of symptom exaggeration (Rogers, 1997, p. 416). This approach entails *a priori* identification of symptom exaggeration unrelated to the MMPI-2 (e.g., sufficient *vs.* insufficient cognitive effort performance), and MMPI-2 profiles or effort performances of over-reporting groups themselves or relative to groups shown to have not exaggerated symptoms are then observed

(e.g., Boone & Lu, 1999; Ross et al., 2004). Response validity meta-analyses have also been conducted (e.g., Nelson et al., 2006; Rogers et al., 1994, 2003; Vickery et al., 2001), which may provide the clinician with a variety of potential moderators to consider, such as gender, criminal *versus* civil litigation context, and type of clinical population examined (e.g., traumatic brain injury, chronic pain, etc.).

Differently, classification tree analysis (CTA) with univariable optimal data analysis (ODA; Yarnold & Soltysik, 2005) is another approach that may be particularly beneficial to the clinician's everyday practice in the examination of psychological *versus* cognitive response validity data. ODA generates decision-making "trees" based upon optimal cut-scores in the anticipation of a dichotomous dependent variable. *Via* inspection of these trees, the clinician is provided a template by which individual cases may be classified according to ODA cut-scores, and then conclude whether individual cases resemble one dichotomous outcome over another. Millis and others (1998) used ODA in the confirmation of neuropsychological test performances in a SG group with histories of mild head injury and a comparison group with histories of moderate and severe TBI. However, we are not aware of any studies to date that have used ODA in the classification of cognitive effort on the basis of MMPI-2 profiles.

In the current study, ODA methodology was applied to better clarify whether MMPI-2 validity scales can improve clinical decision-making with regard to cognitive effort (i.e., IE *vs.* SE). IE and SE status was established as the class variable from which a decision-making "tree" would be grown based upon MMPI-2 scores in a large group of SG and NSG participants. This methodology seems particularly useful to clinicians because it: (1) closely mimics the diagnostic decision-making process and (2) yields optimal MMPI-2 cut scores in the discrimination of cognitive effort that can be applied in future cases. In the context of previous EFA findings (Nelson et al., 2007a), it was anticipated that "somatic/neurotic" scales (e.g., *FBS*, *RBS*, *Md*) would take precedence in the discrimination of IE and SE in the current clinical and forensic sample.

METHOD

Participants

All data included in this manuscript were obtained in compliance with ethical regulations of the institutions at which the data were collected, in compliance with the Helsinki Declaration. Case files were obtained from the archival databases of the third, sixth, and seventh authors in compliance with institutional guidelines, resulting in 307 participants who had completed the MMPI-2 and one or more forced-choice effort measures. One hundred twenty-two of these participants were examined in a separate response validity study (Nelson et al., 2007a). All individuals were referred for neuropsychological evaluation of cognitive complaints.

Of these 307 participants, 198 (64.5%) were evaluated in a secondary gain (SG) context, such as personal injury litigation or in association with an independent medical examination or similar proceedings (e.g., disability, workers' compensation). None of the SG participants were involved in criminal litigation. The remaining 109 (35.5%) had no appreciable secondary gain (NSG). Mean age of the sample was 43.2 years ($SD = 11.9$), with mean education of 13.4 years ($SD = 3.1$). One hundred ninety-three (62.7%) participants were male, and 114 (37.3%) were female. The large majority of SG individuals were referred for evaluation of cognitive complaints associated with traumatic brain injury (111, 56%), compared with only 17% (19 cases) of the NSG group. The remainder of the latter group's referrals were associated with a variety of conditions (e.g., mild head injury, anoxia, pain, ADHD, epilepsy). For the subset of the sample on which data were available ($n = 242$), mean IQ was 100.5 ($SD = 14.4$). There was no significant difference in IQ between the SG and NSG groups ($n = 237$; $t = .754$; $df = 235$; $p = .451$). Application of the conventional 90% rule (e.g., Grote et al., 2000; Inman et al., 1998; Tombaugh, 1996) on one or more of the forced-choice effort measures resulted in 182 individuals being classified as having sufficient effort (SE) and 125 as demonstrating insufficient effort (IE).

Measures

Standard MMPI-2 validity scales were examined in the study, including *L*, *F*, *K*, Back Infrequency (*Fb*), Variable Response Inconsistency Scale (*VRIN*), and True Response Inconsistency Scale (*TRIN*). Additional MMPI-2 validity scales included the *F-K* index (Gough, 1950), Infrequency Psychopathology Scale (*Fp*; Arbisi & Ben-Porath, 1995), the Superlative Scale (*S*; Butcher & Han, 1995), Dissimulation Scale (*Ds*; Gough, 1954), Fake Bad Scale (*FBS*; Lees-Haley et al., 1991), Response Bias Scale (*RBS*; Gervais, 2005), and the Malingered Depression Scale (*Md*; Steffan et al., 2003). In addition to MMPI-2 validity scales, data were available for all clinical scales: Hypochondriasis (*Hs*), Depression (*Dep*), Hysteria (*Hy*), Psychopathic Deviate (*Pd*), Masculinity-Femininity (*Mf*), Paranoia (*Pa*), Psychasthenia (*Pt*), Schizophrenia (*Sc*), Hypomania (*Ma*), and Social Introversion (*Si*). In addition to the MMPI-2, all respondents completed a variety of neuropsychological measures as part of their comprehensive evaluation, and forced-choice effort measures. Effort tests included the Victoria Symptom Validity Test (VSVT; Slick et al., 1995), the Test of Memory Malingered (TOMM; Tombaugh, 1996), the Multi-Digit Memory Test (MDMT; Niccolls & Bolter, 1991), the Word Memory Test (WMT; Green, 2003), and the Letter Memory Test (LMT; Inman et al., 1998).

Analyses and Procedures

CTA using ODA was used to predict cognitive effort status (i.e., IE or SE) in the sample of 307 participants. Although

the concept of ODA had been available previously, the Optimal Data Analysis (ODA) software and general approach is a somewhat new methodology (Bryant, 2005), particularly in the context of neuropsychological research (for a richer, more extensive discussion of this topic, the reader is referred to: Yarnold & Soltysik, 2005). ODA methodology involves identification of a variable and finding the optimal cut point that will accurately classify the greatest number of individuals on a given class variable (for present purposes, use of an MMPI-2 validity scale to predict cognitive effort status). The null hypothesis for this procedure is that the class variable, in this instance IE, cannot be predicted as a linear cutpoint on the continuous (attribute) variable, and the alternate hypothesis is that the class variable can be predicted using this cutpoint (Yarnold, 1996).

CTA *via* ODA consists of constructing a hierarchical decision tree through several different univariate steps. The first step in CTA involves separately conducting univariable ODA analyses for each of the potential predictors or attributes in the model (i.e., MMPI-2 validity and clinical scales), and evaluating their resultant effect strengths. The model yielding the greatest effect strength is then selected. Based on the determined cutpoint for that attribute, some individuals will be classified as SE and some as IE. However, it is expected that some people will be misclassified in both groups. Thus, in an iterative manner, the ODA procedure is repeated to continue to improve classification accuracy using as many of the potential attributes as is necessary. When an attribute no longer improves classification accuracy (as determined by p , effect strength, or number of correct classifications), that branch of the tree is terminated. This procedure is repeated until all branches of the tree have terminated, at which point, a conceptual diagram of the tree can be constructed (Yarnold, 1996). It should be noted that ODA procedures can accommodate both continuous and categorical (i.e., nominal or binary) predictors, without dummy-coding of variables such as gender and race.

Decision-tree methodology involves “growing” a tree from an initial decision point, with branches that denote different courses of action that result from decisions made at lower points in the tree. The tree terminates in a set of outcomes, namely, the assignment of individuals into one of two dichotomous class variables. Most existing tree methodologies do not explicitly maximize classification accuracy as part of their computational algorithm. By contrast, CTA *via* ODA is the only methodology to maximize classification accuracy, by constructing a decision tree that explicitly yields the highest percentage of accurately classified individuals in the sample (Bryant, 2005; Yarnold, 1996).

It is important to note that, although ODA identifies a dichotomous optimal cut score for each predictor in the CTA model, these predictors are not necessarily restricted to having only two levels. Each attribute is analyzed for its predictive power at each potential branch of the tree model, regardless of whether or not the particular attribute has entered the tree at an earlier branch. Thus, an attribute that is dichotomized optimally at one branch can immediately

re-enter the tree model with additional cut scores for either side or both sides of the initial dichotomy, if these additional cut scores for the same attribute contribute to classification accuracy more than cut scores for other attributes (e.g., Donenberg et al., 2003). In this way, nonlinear CTA overcomes the problems of unreliability, low statistical power, and underestimation of effect size that arise from treating continuous attributes solely as binary variables (see MacCallum et al., 2002).

The final CTA model contained predictive attributes that were selected based on the following established procedures. In *growing* the tree, two rules were used: first, we selected the attribute (and accompanying decision rule) with the strongest effect strength (ES) for sensitivity at each node in the classification tree model. ES is an absolute index of effect size for which 0 = performance expected by chance and 100 = perfect classification accuracy. According to Yarnold and Soltysik (2005, p. 61), effect strength values <25% are *weak*, 25–50% are *moderate*, 50–75% are *relatively strong*, 75–90% are *strong*, and >90% are *very strong*. Second, of those attributes, we selected those that provided the highest classification accuracy while remaining stable when submitted to a leave-one-out (LOO) jackknife validity analysis (Lachenbruch, 1967; Yarnold & Soltysik, 2005). More specifically, a jackknife validity analysis examines whether each participant is predicted to be in the IE or SE group using a UniODA model developed from the other $N - 1$ observations (Ostrander et al., 1998). This particular LOO analysis is a measure of expected cross-sample generalizability, and provides information on how likely it is that a model will cross-validate and accurately classify future individuals into those with SE and those without. While the “gold standard” on cross-validation is to re-test the model in a second, independent sample, LOO analysis provides a useful alternative for assessing expected cross-sample generalizability in situations where such data are not available.

After the initial tree was constructed, two rules were used to *prune* the tree. First, we determined the statistical significance of each attribute in the final model by performing a nondirectional Fisher’s exact probability test, with $\alpha = .05$. Second, we used a sequentially rejective Bonferroni procedure to further prune the tree, to ensure an experiment-wise Type I error rate of $p < .05$. More specifically, we used a Sidak step-down adjustment procedure (Yarnold & Soltysik, 2005) to prune nodes from the tree if their type I error exceeded 0.05, controlling for the number of nodes in the final tree model. This latter procedure was used to craft the most parsimonious model while not capitalizing on chance.

RESULTS

After pruning, the final tree (Figure 1) contained two attributes: MMPI-2 Response Bias Scale (*RBS*) and MMPI-2 Hysteria Scale (*Hy*). By definition, sensitivity refers to the number of true positives/(true positives + false negatives), while specificity refers to the number of true negatives/(true

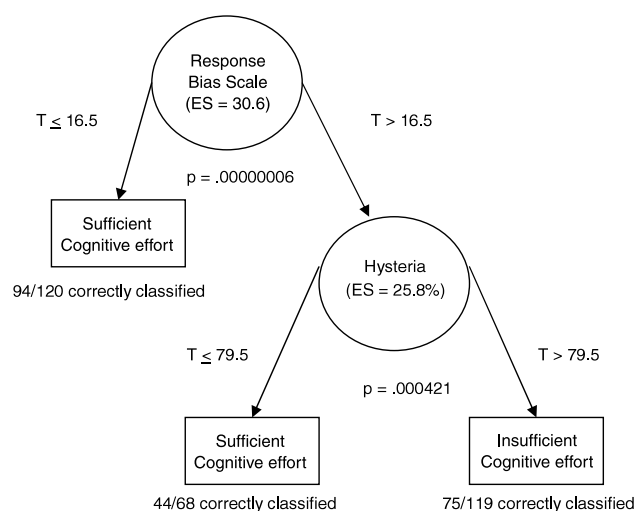


Fig. 1. Diagram of the hierarchically optimal classification tree model for predicting sufficient (0) versus insufficient (1) cognitive effort among adult outpatients presenting for neuropsychological evaluation using all 26 predictors and adopting a sequentially rejective Bonferroni adjustment ($p < .05$) to prune the tree model ($n = 307$). In this figure, circles represent nodes (or decision points) containing each predictive attribute and its effect strength (ES, in parentheses), arrows represent branches (or predictive pathways), and rectangles represent prediction endpoints (or final classifications). Numbers (probabilities) centered beneath nodes are the generalized p value for each node, based on nondirectional Fisher's exact test. Numbers (inequalities) beside arrows indicate the value of the cut-point for optimally classifying observations into categories for each node (decision rule). Fractions beneath each prediction endpoint represent the number of correct classifications at the endpoint (numerator) and total number of observations at the endpoint (denominator). Numbers in parentheses next to fractions are the predictive value for each endpoint (or percentage of predicted classifications into the given category that were correct).

negatives + false positives) (Table 2). Thus, an IE participant predicted as such was a true positive, an IE participant predicted as SE was a false negative, an SE participant predicted as SE was a true negative, and an SE participant predicted as IE was a false positive. By these guidelines, *RBS* and *Hy* accurately classified 69.4% of the sample in terms of whether they gave insufficient effort (sensitivity = 60%) or not (specificity = 75.8%). The overall effect strengths for sensitivity and specificity were 20.0% and 51.6%, respectively, where zero is the performance level expected by chance and 100 is perfect classification accuracy. Thus, the current CTA model has a weak overall sensitivity but relatively strong specificity in predicting whether or not an individual has given insufficient effort on cognitive tests.

While sensitivity and specificity are useful, of arguably greater utility to psychologists are the indices of positive and negative predictive validity, where positive predictive validity is the number of true positives/(true positives + false positives) and negative predictive validity is the number of true negatives/(true negatives + false negatives). In the current analyses, positive predictive validity was 63.0%,

while negative predictive validity was 73.4%. Overall effect strength of the predictive value was 36.4% which, by Yarnold and Soltysik's (2005) standards, constitutes a moderate effect strength.

In examining the actual CTA more closely (Figure 1), at the top of the tree model is the novel MMPI-2 Response Bias Scale (*RBS*), forming the primary node in the hierarchically optimal tree model (i.e., it is the predictive attribute with the strongest effect size for the total sample). Thus, *RBS* is essential in determining whether or not someone is likely to have given IE on effort tests. If an individual provides an *RBS* score less than or equal to 16.5, then that individual is determined with 78.3% accuracy as having given good effort. These individuals comprise 31% of the sample. If, however, *RBS* is greater than 16.5, then the MMPI-2 Hysteria (*Hy*) scale enters the model. Those who score $T > 79.5$ on the *Hy* scale are classified with 75.0% accuracy as having given IE. Thus, someone who endorses many unusual cognitive complaints (such as those tapped by *RBS*) in addition to vague and nonspecific somatic concerns (i.e., items on *Hy*), is likely to have given insufficient effort on cognitive testing. These individuals represent 24% of the sample. By contrast, those who score $T \leq 79.5$ on the *Hy* scale are classified with 64.7% accuracy as having given SE on cognitive tests, representing 14% of the entire sample.

As *RBS* is a newer and lesser-known scale that may not be in common usage among neuropsychologists, we re-conducted ODA analyses with all MMPI-2 variables, except for *RBS*. The same procedures, including growing and pruning rules, were used as outlined in the prior set of analyses. When *RBS* was removed from the model and after pruning, *Hy* took precedence in classifying cognitive effort status (see Figure 2), with *Pa*, *Fb*, *K*, and *Fp* as subsequent predictive attributes. These five scales accurately classified 71.0% of the sample in terms of whether they gave IE (sensitivity = 58.5%) or not (specificity = 79.4%). The overall effect strengths for sensitivity and specificity were 17.0% (weak) and 58.8% (relatively strong), respectively. By contrast, positive predictive validity was 65.7%, while negative predictive validity was 73.9%. Overall effect strength of the predictive value was 39.6%, which constitutes a moderate effect strength.

Of note is that, after eliminating *RBS* from the analysis, *Hy* now assumes the primary node in the hierarchically optimal tree model, crucial in determining whether or not someone is likely to have given IE on cognitive tests (Figure 2). Again, an optimal cutting score of $T = 79.5$ on *Hy* serves to define the classification tree. On the left branch (*Hy*: $T \leq 79.5$), an individual's score on the *Pa* scale defines their effort status. That is, when $T = 54.5-60$ on the *Pa* scale, that person was classified with 57.1% accuracy as having given IE. When the elevation on *Pa* was $T \leq 54.5$ or >60 , that person was classified as having given SE.

By contrast, when the elevation on *Hy* is $T > 79.5$, two further decision-points presented themselves, based on the degree of *Fb* elevation. First a "defensiveness" branch of the tree emerged with $T \leq 68.5$ on *Fb*. If $T \leq 55$ on *K*, that

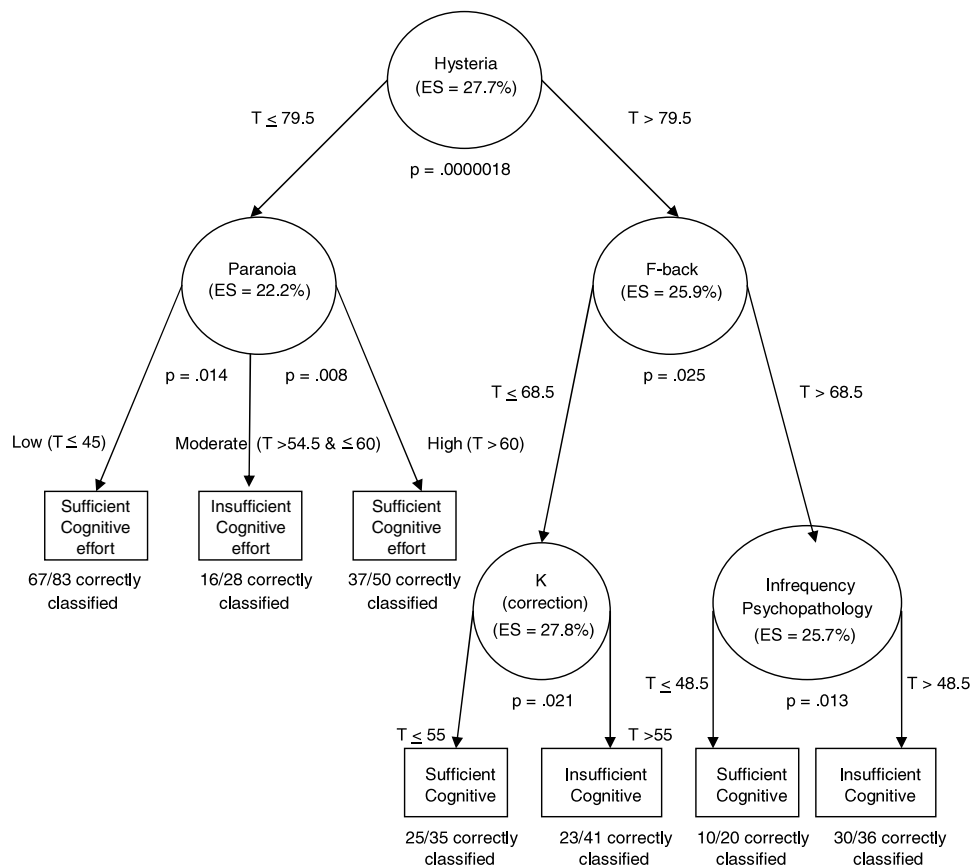


Fig. 2. Diagram of the hierarchically optimal classification tree model for predicting sufficient (0) versus insufficient (1) cognitive effort among adult outpatients presenting for neuropsychological evaluation, using all predictors except RBS ($n = 293$). All attributes in this tree model were statistically significant at $p < .05$, regardless of whether or not the Bonferroni adjustment was imposed.

individual was classified with 71.4% accuracy as having given SE, whereas if $T > 55$ on K , there was 56.1% classification accuracy of IE. This left a second, “eager-to-overreport” branch of the tree, with $T > 68.5$ on Fb . If $T \leq 48.5$ on Fp , that individual was classified with 50.0% accuracy as having given SE, whereas if $T > 48.5$ on K , there was 83.3% classification accuracy of IE. Looking at relative classification accuracies for these two branches together, it appears that ODA had most utility in predicting effort status in those who scored either consistently low on the scales of interest (i.e., low Fb , low K , SE = 71.4% accuracy) or consistently high (i.e., high Fb , high Fp , IE = 83.3% accuracy). To illustrate the clinical accuracy and statistical results of the tree results in a manner more familiar to clinicians, Tables 1 and 2 provide detailed findings showing the effects of including and omitting RBS .

In contrast to a model’s sensitivity (i.e., the probability that a person who actually belongs in a particular category will be correctly classified as being in that category), a model’s predictive value (i.e., the probability that a person classified as being in a particular category actually belongs in that category) varies as a function of the actual base rate in the population and the model’s rate of incorrect classifications. For this reason, it is important to assess the impact

of different population base-rates on the utility of a model with a given rate of incorrect classifications (Ostrander et al., 1998; Wainer, 1991). A classification model is *efficient* if it provides a predictive value greater than the population base-rate (Meehl & Rosen, 1955). For example, a model that classifies observations with a predictive value equal to the population base-rate performs no better than chance, and clinicians would be better off not administering the set of measures included in the model and instead simply guessing that the chance of any given observation belonging in the category is equal to the base rate. In this case, the model would be said to lack efficiency.

We computed measures of efficiency (Meehl & Rosen, 1955; Yarnold & Soltysik, 2005, p. 60–61) for both positive predictive value (in classifying insufficient cognitive effort) and negative predictive value (in classifying sufficient cognitive effort) for the Bonferroni-adjusted CTA models with and without the RBS scale. Figure 3 displays these estimates of efficiency for both models. As seen in this figure, both CTA models perform best compared to chance in classifying observations in either category for population base-rates between 0.3 and 0.5. (Note that the base-rate of insufficient cognitive effort for the present sample was roughly 0.4.)

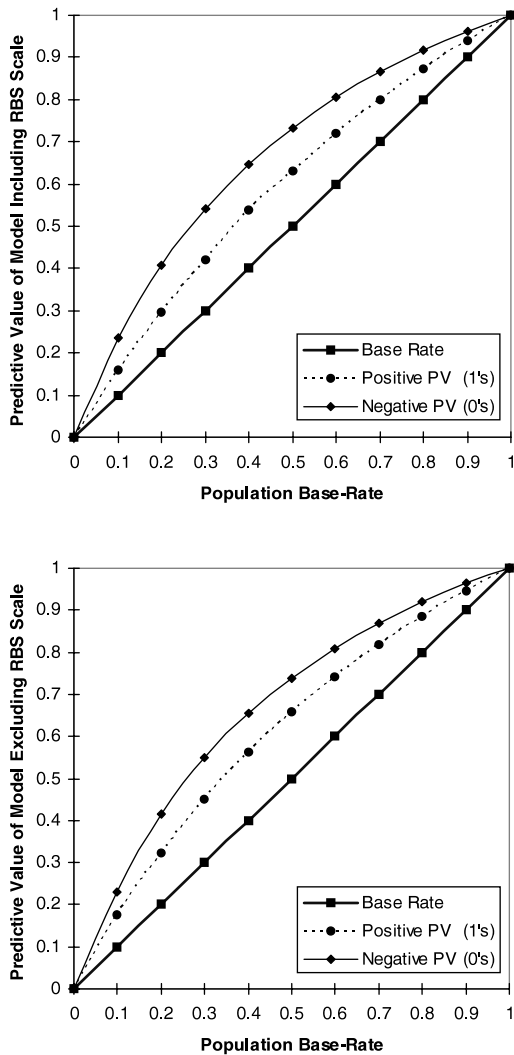


Fig. 3. Estimates of classification efficiency for both positive predictive value (in classifying insufficient cognitive effort) and negative predictive value (in classifying sufficient cognitive effort) as a function of different population base-rates, for the Bonferroni-adjusted CTA model including the RBS scale (top graph) and the Bonferroni-adjusted CTA model excluding the RBS scale (bottom graph).

DISCUSSION

The current study is unique in its use of classification tree methodology; no other studies to date have examined psychological and cognitive response validity measures with this approach. Derivation of an optimal CTA model allows for a greater degree of sophistication in classifying individuals from known groups, rather than applying a “one-size-fits-all” regression model to an entire sample of individuals. That is, optimal CTA methodology more closely mimics the differential diagnostic decision-making process more commonly used in clinical practice, enhancing the practical utility and ecological validity of the MMPI-2 in the forensic context. Thus, rather than examining the same scales for all individuals, different scales with different cut scores are

examined in concert, leading to a more tailored approach to classification. All MMPI-2 validity and clinical scales were allowed to compete for inclusion in the model, which after “pruning” of the initial “tree,” resulted in only two variables, which provides the clinician with a parsimonious understanding of the findings. This is an advantage over traditional regression models, where using only the minimum number of variables to create a classification tree that is robust enough to be expected to replicate across samples circumvents capitalization on chance or idiosyncratic response patterns.

The purpose of this study was to investigate, through use of a unique statistical approach (ODA), the degree to which MMPI-2 validity and clinical scales predict the result of cognitive effort tests (IE/SE) in a large group of NSG and SG participants. ODA is a novel yet powerful methodology that mimics the diagnostic decision-making process in the clinical context. The resultant analyses identified optimal classification “branches” of MMPI-2 validity and clinical scales at different cut scores, providing MMPI-2 scores that may be expected to identify sufficient *versus* insufficient cognitive effort. Consistent with the expectation that over-reporting of somatic symptoms may be more pertinent to cognitive effort than over-reporting of psychotic/rarely endorsed symptoms (Nelson et al., 2007a), *RBS* took precedence in classifying cognitive effort in the current sample. This may relate to the original rationale behind *RBS* development, whereby MMPI-2 test items were chosen on the basis of discrimination of forced-choice effort (*WMT*) performance. In clinically relevant terms, results of the initial ODA model (see Figure 1) suggest that: (1) low *RBS* scores (<16.5) tend toward SE, (2) high *RBS* scores (>16.5) and high *Hy* scores (>79.5) tend toward IE, and (3) high *RBS* (>16.5) and low *Hy* scores (<79.5) tend toward SE.

Additionally, as *RBS* is a little known and relatively new scale, we decided that a separate ODA with *RBS* excluded might be of additional benefit to the clinician. Somewhat surprisingly, a clinical scale (*Hy*), rather than a validity scale, took precedence when *RBS* was removed from the analysis. Based upon the item content of *Hy*, results further support the notion that “somatic” symptoms have a unique relationship with cognitive effort in the current clinical and civil forensic sample, and the magnitude of the optimal *Hy* cutoff ($T > 79.5$) seems very much relevant to clinical practice and the assessment of psychological response validity. Graham (2000, p. 69) suggests that an elevation of this magnitude is consistent with an individual who “reacts to stress and avoids responsibility by developing physical symptoms”. In this context, present findings suggest that demonstration of insufficient cognitive effort may represent an additional method of avoiding responsibility in litigants and claimants who over-report somatic symptoms. Likely, this is related to the fact that the SG individuals referred for neuropsychological evaluation most often present with a host of somatic complaints that may be expressed particularly on *Hy* and not necessarily validity scale elevations (Lanyon & Almer, 2002; Lees-Haley, 1997). The concep-

Table 1. Overall cross-classification tables for the two ODA tree models (i.e., 26 MMPI-2 indices with and without RBS), predicting whether individuals exerted sufficient (0) or insufficient (1) cognitive effort

Attributes included in the analysis	Bonferroni adjustment?	Actual level of cognitive effort	Predicted level of cognitive effort		Classification accuracy	
			Sufficient effort (0)	Insufficient effort (1)	Row total	(PAC)
All 26 MMPI scales, sex, age, & education	No (<i>N</i> = 276)	0	88	77	165	(53.3%)
		1	36	75		
	Yes (<i>N</i> = 307)	(PV)	(71.0%)	(49.3%)	182	(75.8%)
		0	138	44		
25 MMPI scales, (omitting RBS), sex age, & education	No (<i>N</i> = 293)	0	139	36	175	(79.4%)
		1	49	69		
	Yes (<i>N</i> = 293)	(PV)	(73.9%)	(65.7%)	175	(79.4%)
		0	139	36		
		(PV)	(73.9%)	(65.7%)	<i>p</i> = .048 × 10 ⁻⁹	

Note. When omitting RBS, the same attributes in the tree model were statistically significant at nondirectional *p* < .05, regardless of whether or not a Bonferroni adjustment (i.e., sequentially rejective Sidak procedure, Bonferroni *p* < .05) was adopted. PAC = percentage of classification accuracy, or the proportion of observations in each level of the class variable that were correctly classified; PV = predictive value, or the proportion of predicted classifications that were actually correct; *N* = the total number of observations classified by the set of predictors in the particular tree model, excluding observations with missing values on the specific set of attributes used to classify them. *p* = nondirectional Fisher's exact probability.

Table 2. Classification performance statistics for the two ODA tree models (i.e., 26 MMPI-2 indices with and without RBS) predicting whether individuals exerted sufficient (0) or insufficient (1) cognitive effort

Performance index	Performance parameter (%)			
	MMPI-2 including RBS		MMPI-2 without RBS	
	Generalized <i>p</i> < .05 (<i>N</i> = 276)	Bonferroni <i>p</i> < .05 (<i>N</i> = 307)	Generalized <i>p</i> < .05 (<i>N</i> = 293)	Bonferroni <i>p</i> < .05 (<i>N</i> = 293)
Overall classification accuracy	163/276 (59.1%)	213/307 (69.4%)	208/293 (71.0%)	208/293 (71.0%)
Sensitivity (IE)	75/111 (67.6%)	75/125 (60.0%)	69/118 (58.5%)	69/118 (58.5%)
Specificity (SE)	88/165 (53.3%)	138/182 (75.8%)	139/175 (79.4%)	139/175 (79.4%)
ES for Sensitivity	35.2%	20.0%	17.0%	17.0%
ES for Specificity	6.6%	51.6%	58.8%	58.8%
PV (IE)	75/152 (49.3%)	75/119 (63.0%)	69/105 (65.7%)	69/105 (65.7%)
PV (SE)	88/124 (71.0%)	138/188 (73.4%)	139/188 (73.9%)	139/188 (73.9%)
ES for Predictive Value	20.3%	36.4%	39.6%	39.6%

Note. The first CTA model included 23 MMPI subscales, sex, age, and years of education. The second CTA model excluded RBS from the analysis. Omitting RBS, the same attributes were retained in the tree model as being statistically significant at nondirectional *p* < .05, regardless of whether or not a Bonferroni adjustment (i.e., sequentially rejective Sidak procedure, Bonferroni *p* < .05) was adopted. Overall classification accuracy = the total number of actual 1s and 0s that were correctly classified. Sensitivity (insufficient effort) = percentage of classification accuracy for observations with true values of 1 on the class variable. Specificity (sufficient effort) = percentage of classification accuracy for observations with true values of 0 on the class variable. *N* = the total number of observations classified by the set of predictors in the particular tree model, excluding observations with missing values on the specific set of attributes used to classify them.

tual role of *Hy* as the preliminary point of decision-making (i.e., after excluding *RBS*) is therefore of clinical interest. Based upon the clinician's evaluation of *Hy*, determination of which pattern of potential symptom exaggeration (somatic or psychiatric) can more likely be made. With significantly elevated *Hy*, the clinician should be especially mindful of the possibility of somatic malingering, which would consist of concurrent elevations of *Hs*, *Hy*, and *FBS* (Larrabee, 1998). Without significantly elevated *Hy*, psychiatric symptom exaggeration may be of greater likelihood. It is of note that, despite being entered as a predictor in both ODA models, the *Fake Bad Scale (FBS)* failed to enter the model as a significant predictor of effort status regardless of whether the *RBS* was included or not. This is of interest given the fact that the *FBS*, like the *RBS*, was designed for use in the civil litigation context.

Two findings from Figure 2 seem particularly relevant to clinical practice. First, high *Hy* scores, matched with elevations on two F-family scales (*Fb*, *Fp*) tend toward IE. In other words, individuals who demonstrate excessive somatic symptoms, and who simultaneously endorse some degree of psychotic/rarely-endorsed symptoms are also more likely to show IE. Second, individuals who endorse fewer somatic symptoms, and who also endorse minimal psychotic symptoms (*Pa*) are more likely to show SE. Taken together, IE is more likely associated with endorsement of both "somatic" and "psychotic" symptoms, while SE is more likely associated with lesser endorsement of these same symptoms.

Other findings in Figure 2 are more difficult to grasp in terms of clinical utility. One possible explanation behind the complexity of Figure 2 is the extent of effort variance that somatic symptoms account for in the initial stages of the model. That is, after accounting for the somatic symptoms endorsed on *RBS* and *Hy*, it is possible that psychological validity and clinical scales play little role in accurate detection of cognitive effort. Indeed, overall, ODA classification rates were at times satisfactory (as high as 83.3%), while other classification rates were unacceptably low (50.0%). Overall strengths in the ODA model ranged from low (i.e., <25%) to moderate in magnitude (i.e., 25–30%), suggesting that MMPI-2 validity and clinical scales cannot be consistently expected to predict IE versus SE in the current sample. In other words, variable classifications and moderate effect strengths suggest that there is *not* likely to be a consistent, one-to-one correspondence among MMPI-2 validity and clinical scales and measures of cognitive effort. Such a viewpoint seems consistent with the results of a recent EFA (Nelson et al., 2007a), which suggested that while somatic over-reporting appears to have a greater relationship with cognitive effort relative to other scales (e.g., F-family), correlations were modest nonetheless. As such, we emphasize that clinicians should never abandon use of cognitive effort tests in favor of MMPI-2 findings. Even with use of a sophisticated and clinician-friendly statistical methodology such as ODA, one cannot assume that MMPI-2 validity and clinical scales will effectively predict IE versus SE in the forensic context.

Literature on the interaction of cognitive and psychological effort variables is sparse to the degree that *a priori* theorizing about the relationships between such variables would be quite difficult. As such, we view the current findings in the context of *theory building* and as such are illustrative rather than prescriptive regarding how ODA methodology can be used in the forensic context. We would certainly hope that our findings would provoke further research in this area, using the current results in a more theoretically driven, *theory-testing* manner. In replicating our findings, we make several recommendations to address limitations of the current research. First, having two large independent samples provides the best means to ascertain whether a model created for one sample will generalize to a second sample. Having two samples will also allow for mixed group validation (MGV; Dawes & Meehl, 1966; Frederick, 2000), where it is not necessary to know beforehand the exact proportions of individuals who are in SG and NSG contexts. Third, where possible, having uniform neuropsychological data across participants would allow for quantification of the effect of effort status on test scores and how this differs depending on the SG/NSG context.

In summary, based upon variable classification accuracies generated, current findings provide further evidence that both types of response validity measures (i.e., psychological and cognitive) do not directly correspond with one another, and both are necessary to obtain an accurate understanding of response validity for an individual patient or litigant.

ACKNOWLEDGMENTS

The authors have no financial or other relationship that would constitute a conflict of interest related to the research represented in this study. There was no commercial or other financial support for this project.

REFERENCES

- Arbisi, P.A. & Ben-Porath, Y.S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The infrequency psychopathology scale *F(p)*. *Psychological Assessment*, 7, 424–431.
- Bagby, R.M., Nicholson, R.A., Buis, T., & Bacchiochi, J.R. (2000). Can the MMPI-2 validity scales detect depression feigned by experts? *Assessment*, 7, 55–62.
- Bianchini, K.J., Mathias, C.W., & Greve, K.W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist*, 15, 19–45.
- Boone, K.B. & Lu, P.H. (1999). Impact of somatoform symptomatology on credibility of cognitive performance. *The Clinical Neuropsychologist*, 13, 414–419.
- Bryant, F.B. (2005). How to make the best of your data. *PsycCRITIQUES*, 50.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Laemmer, B. (1989). *Manual for Administration and Scoring of the Minnesota Multiphasic Personality Inventory* (2nd ed.). Minneapolis, MN: University of Minnesota Press.
- Butcher, J.N. & Han, K. (1995). Development of an MMPI-2 scale to assess the presentation of self in a superlative manner: The S

- scale. In J.N. Butcher & C.D. Spielberger (Eds.), *Advances in Personality Assessment, Vol. 10*. (pp. 25–50). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dawes, R.M. & Meehl, P.E. (1966). Mixed-group validation: A method for determining the validity of diagnostic signs without using criterion groups. *Psychological Bulletin*, *66*, 63–67.
- Dearth, C.S., Berry, T.R., Vickery, C.D., Vagnini, V.L., Baser, R.E., Orey, S.A., & Cragar, D.E. (2005). Detection of feigned head injury symptoms on the MMPI-2 in head injured patients and community controls. *Archives of Clinical Neuropsychology*, *20*, 95–110.
- Donenberg, G.R., Bryant, F.B., Emerson, E., Wilson, H.W., & Pasch, K.E. (2003). Tracing the roots of early sexual debut among adolescents in psychiatric care. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 594–608.
- Frederick, R.I. (2000). Mixed group validation: A method to address limitations of criterion group validation in research on malingering detection. *Behavioral Sciences and the Law*, *18*, 693–718.
- Gervais, R. (2005, April). *Development of an Empirically Derived Response Bias Scale for the MMPI-2*. Paper presented at the Annual MMPI-2 Symposium and Workshops, Ft. Lauderdale, FL.
- Gough, H.G. (1950). The F minus K dissimulation index for the MMPI. *Journal of Consulting Psychology*, *14*, 408–413.
- Gough, H.G. (1954). Some common misconceptions about neuroticism. *Journal of Consulting Psychology*, *18*, 287–292.
- Graham, J.R. (2000). *MMPI-2: Assessing Personality and Psychopathology*. Third Edition. New York: Oxford University Press.
- Green, P. (2003). *Word Memory Test for Windows: User's Manual and Program*. Edmonton, Alberta, Canada: Author. (Revised 2005).
- Greiffenstein, M.F., Gola, T., & Baker, J.W. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist*, *9*, 230–240.
- Grote, C.L., Kooker, E.K., Nyenhuis, D.L., Smith, C.A., & Mattingly, M.L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical and Experimental Neuropsychology*, *22*, 709–719.
- Hiscock, M. & Hiscock, C.K. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, *11*, 967–974.
- Inman, T.H., Vickery, C.D., Berry, D.T.R., Lamb, D.G., Edwards, C.L., & Smith, G.T. (1998). Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The Letter Memory Test. *Psychological Assessment*, *10*, 128–139.
- Lachenbruch, P.A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, *23*, 639–645.
- Lanyon, R.I. & Almer, E.R. (2002). Characteristics of compensable disability patients who choose to litigate. *Journal of the American Academy of Psychiatry and the Law*, *30*, 400–404.
- Larrabee, G.J. (1998). Somatic malingering on the MMMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist*, *12*, 179–188.
- Larrabee, G.J. (2003). Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist*, *17*, 54–68.
- Lees-Haley, P.R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology*, *53*, 745–755.
- 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*, 208–210.
- MacCallum, R.C., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40.
- Meehl, P.E. & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*, 194–216.
- Millis, S.R., Ross, S.R., & Ricker, J.H. (1998). Detection of incomplete effort on the Wechsler Adult Intelligence Scale-Revised: A cross-validation. *Journal of Clinical and Experimental Neuropsychology*, *20*, 167–173.
- Mittenberg, W.M., Patton, C., Canyock, E.M., & Condit, D.C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*, 1094–1102.
- Nelson, N.W., Sweet, J.J., Berry, D.T.R., Bryant, F.B., & Granacher, R.P. (2007a). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society*, *13*, 440–449.
- Nelson, N.W., Sweet, J.J., & Demakis, G. (2006). Meta-analysis of the MMPI-2 Fake Bad Scale: Utility in forensic practice. *The Clinical Neuropsychologist*, *20*, 39–58.
- Nelson, N.W., Sweet, J.J., & Heilbronner, R. (2007b). Examination of the new MMPI-2 Response Bias Scale (Gervais): Relationship with MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology*, *29*, 67–72.
- Niccolls, R. & Bolter, J. (1991). *Multi-Digit Memory Test (Computer Version)*. Los Angeles, CA: Western Psychological Services.
- Ostrander, R., Weinfurt, K.P., Yarnold, P.R., & August, G.J. (1998). Diagnosing attention deficit disorders using the BASC and the CBCL: Test and construct validity analyses using optimal discriminant classification trees. *Journal of Consulting and Clinical Psychology*, *66*, 660–672.
- Pope, K., Butcher, J., & Seelen, J. (2000). *The MMPI, MMPI-2, & MMPI-A in Court* (2nd ed.). Washington, DC: American Psychological Association.
- Posthuma, A.B. & Harper, J.F. (1998). Comparison of MMPI-2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice*, *29*, 437–443.
- Rogers, R. (Ed.) (1997). *Clinical Assessment of Malingering and Deception (2nd Edition)*. New York: Guilford.
- Rogers, R., Sewell, K.W., Martin, M.A., & Vitacco, M.J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*, 160–177.
- Rogers, R., Sewell, K.W., & Salekin, R.T. (1994). A meta-analysis of malingering on the MMPI-2. *Assessment*, *1*, 227–237.
- Rogers, R., Sewell, K.W., & Ustad, L.L. (1995). Feigning among chronic outpatients on the MMPI-2: A systematic examination of fake-bad indicators. *Assessment*, *2*, 81–89.
- Ross, S.R., Millis, S.R., Krukowski, R.A., Putnam, S.H., & Adams, K.M. (2004). Detecting incomplete effort on the MMPI-2: An examination of the Fake Bad Scale in mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *26*, 115–124.
- Slick, D.J., Hopp, G.A., & Strauss, E.H. (1995). *The Victoria Symptom Validity Test*. Odessa, FL: PAR.

- Slick, D.J., Sherman, E.M.S., & Iverson, G.L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.
- Steffan, J.S., Clopton, J.R., & Morgan, R.D. (2003). An MMPI-2 scale to detect malingered depression (*Md* Scale). *Assessment*, *10*, 382–392.
- Tombaugh, T.N. (1996). *Test of Memory Malingering*. Toronto, Ontario: MultiHealth Systems.
- Vickery, C.D., Berry, D.T.R., Inman, T.H., Harris, M.J., & Orey, S.A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*, 45–73.
- Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. *Psychological Bulletin*, *109*, 147–151.
- Yarnold, P.R. (1996). Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement*, *66*, 656–667.
- Yarnold, P.R. & Soltysik, R.C. (2005). *Optimal Data Analysis: A Guidebook with Software for Windows*. Washington, DC: American Psychological Association.