

The latent structure of cognitive symptom exaggeration on the Victoria Symptom Validity Test

Thomas W. Frazier^{a,b,*}, Eric A. Youngstrom^c, Richard I. Naugle^b,
Kathryn A. Haggerty^b, Robyn M. Busch^b

^a Section of Behavioral Medicine, The Cleveland Clinic, United States

^b Section of Neuropsychology, The Cleveland Clinic, United States

^c Department of Psychology, University of North Carolina at Chapel Hill, United States

Accepted 27 December 2006

Abstract

Previous studies have focused on the ability of cognitive symptom validity tests to identify simulated malingering or distinguish between clinical samples of individuals at low or high risk of cognitive symptom exaggeration. However, no published studies have examined the latent structure of negative response bias on cognitive tests: measures of cognitive symptom exaggeration may evaluate a continuum of poor effort/invalid responding or a dichotomy of adequate versus inadequate effort. The present study examined whether Victoria Symptom Validity Test (VSVT) indices evaluate a latent dimension or category of response distortion. The VSVT and personality data were obtained from 300 individuals who participated in neuropsychological evaluations as part of standard clinical care. Results indicated that VSVT accuracy scores measure a latent category of inadequate/adequate effort. Individuals classified as taxon members showed significantly poorer performance IQ and memory relative to individuals not classified as exhibiting distortion. The base rate of the identified cognitive symptom exaggeration taxon was estimated to be approximately .13–.14 in the present sample. Likelihood ratios are presented to assist clinical detection of individuals exhibiting the category of cognitive symptom exaggeration.

© 2007 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.

Keywords: Malingering; Taxometric; Victoria Symptom Validity Test; Response bias; Cognitive symptom exaggeration

Detection of negative response bias¹ has long been a concern in personality assessment. Both the MMPI-2 and the Personality Assessment Inventory (PAI; Morey, 1991) include measures of positive and negative impression management or response bias (Greene, 1997; Morey, 2003; Webb, 1999). These measures are quite useful for the detection of psychopathological symptom exaggeration (Arbisi & Ben-Porath, 1998; Lewis, Simcox, & Berry, 2002; Rogers,

* Corresponding author at: Section of Behavioral Medicine (CR11), 2801 Martin Luther King Jr. Drive, Cleveland, OH 44104, United States. Tel.: +1 216 721 1285; fax: +1 216 721 1389.

E-mail address: fraziet2@ccf.org (T.W. Frazier).

¹ The term negative response bias is used generally to refer to any form of symptom distortion. The terms negative response bias on cognitive tasks, cognitive symptom exaggeration, poor effort, cognitive dissimulation, symptom validity, and feigning are used interchangeably in this literature to refer to symptom exaggeration on cognitive measures. However, some of these terms are confusing and may refer to slightly different aspects or types of response bias. We prefer the terms cognitive symptom exaggeration and psychopathological symptom exaggeration, since these terms more accurately characterize different classes and directions of response distortion. Malingering, although frequently used in the literature to refer to symptom exaggeration, is a specific diagnosis that requires additional criteria above and beyond symptom exaggeration.

Sewell, Cruise, Wang, & Ustad, 1998; Storm & Graham, 2000), but they have limited utility for the detection of poor effort or cognitive symptom exaggeration (Berry & Butcher, 1998). For example, Slick and colleagues (Slick, Hopp, Strauss, & Spellacy, 1996) demonstrated modest associations between measures of feigned memory impairment and MMPI-2 validity scales, suggesting that cognitive and psychopathological symptom exaggeration only partially overlap.

As a result, clinicians and researchers have been paying increasing attention to the development and validation of measures designed to detect negative response bias on cognitive tests (for review and discussion of this issue see Bender & Rogers, 2004; Franzen & Iverson, 1998; Nies & Sweet, 1994; Rosenfeld, Sands, & Van Gorp, 2000). Numerous studies have examined the ability of measures of cognitive symptom exaggeration/validity to detect simulated malingerers or distinguish between clinical groups expected to differ substantially in rates of malingering (i.e., mild TBI litigants versus non-litigants) (for recent examples see Doss, Chelune, & Naugle, 1999; Grote et al., 2000; Langeluddecke & Lucas, 2003, 2004; Langeluddecke & Lucas, 2004; Thompson, 2002). However, to the authors' knowledge, no studies have examined the latent structure of cognitive symptom exaggeration.

Initially, measures of cognitive symptom exaggeration were quite simple, yielding a single score meant to evaluate the likelihood of feigning cognitive impairment. Over the last decade, existing measures have become increasingly differentiated. This has been at least partly driven by the observation that individuals may attempt to feign or exaggerate specific types of cognitive impairment. Thus, individuals may perform poorly on only some types of tests, items, or response formats. Feigning of more specific cognitive impairments has resulted in two general types of symptom validity measures. The first type measures non-memory dissimulation, an example being the Validity Indicator Profile (Frederick, 2002), and the second type is an elaboration of early memory malingering measures, involving examination of an individual's pattern of performance. The Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997) is a prototypical example of the latter measures (for studies establishing VSVT validity see Grote et al., 2000; Slick et al., 2003). The VSVT includes multiple indices designed to assess accuracy and response latency using several manipulations, including variation of apparent item difficulty and stimulus-response delay. Thus, the VSVT may capture multiple, distinct dimensions or types of cognitive symptom exaggeration.

1. Categorical versus continuous

Recent work has suggested that psychopathological symptom exaggeration, as assessed by the MMPI-2 F and F(p) scales, has a categorical latent structure (Strong, Greene, & Schinka, 2000). However, we are unaware of any published work that has examined the dimensional versus categorical structure of cognitive symptom exaggeration. One might hypothesize the existence of a distinct category of cognitive symptom exaggeration based upon the notion that some individuals present for evaluation in order to obtain personal gain and make a conscious decision to present themselves unfavorably. Alternatively, a quick glance at the major clinical neuropsychology listserv would indicate that many clinicians view cognitive symptom exaggeration as a continuum resulting from varying levels of poor effort. Distinguishing between these two views has tremendous import for the interpretation of neuropsychological test results. Specifically, identification of the true latent structure of cognitive symptom validity is important for inferring the meaning of validity and cognitive test results, improving discrimination of individuals who are exaggerating cognitive problems from those who are not feigning, and establishing base rates of dissimulation in clinical settings.

Previous research concerning malingering of cognitive deficit has focused primarily on the identification of cutoffs that maximize the predictive accuracy of symptom validity measures. This work has been complicated by lack of attention to the influence of base rates on the efficiency of cut scores (Meehl & Rosen, 1955; Rosenfeld et al., 2000), reliance on samples of simulated malingerers who may not be representative of real-world feigning or exaggeration, and the proliferation of multiple malingering indicators and cut scores. For example, Grote et al. (2000) reported that a cut-score of <90% correct on the VSVT hard items had adequate sensitivity (64%); however, this cut score also resulted in 7% of non-compensation seeking patients being classified as giving sub-optimal effort. This and other cut scores have limited utility because they are typically based on comparisons between compensation seeking (or litigating) and non-compensation seeking (or non-litigating) patients whose cognitive symptom exaggeration status is *assumed*. Unfortunately, the efficiency of the cut score derived in these cases is directly dependent on whether *all* of the individuals in each of these groups are behaving as expected, an unlikely occurrence. This approach is also less helpful clinically because scores close to the cut score are treated as being equally probable of feigning as scores well below the cut score.

Alternatively, other studies have focused on determining cut scores identifying individuals who are choosing the incorrect answer at highly improbable levels based upon binomial probability theory. This approach suffers from a lack of sensitivity, because few individuals score low enough to meet these stringent cut scores (Slick et al., 2003). However, this approach is still frequently used clinically, particularly in legal settings, because of the straightforward, probability-based foundation of interpretations.

Finally, Slick, Sherman, & Iverson (1999) proposed a set of criteria for identifying individuals who are malingering neurocognitive dysfunction (MND). According to these criteria, patients can fall into one of four categories: definite MND, probable MND, possible MND, or not meeting criteria for MND. While useful for research, graded categories are not clinically useful because clinicians' typically must make a dichotomous decision to interpret or not interpret cognitive test results. Additionally, several of the criteria have unknown inter-rater reliability or may be outside of the scope of knowledge of the practitioner (e.g., presence of substantial external incentive). Thus, there remains no "gold standard" criterion or classification strategy that is clinically practical and not arbitrarily defined based upon untested assumptions.

Similarly, estimation of the true base rate of cognitive dissimulation remains guesswork based upon estimates derived from evaluation of high-risk groups. Previous research has suggested relatively high estimates of feigned cognitive impairment in typical outpatient settings (15%; Rogers, Harrell, & Liff, 1993; Rogers, Salekin, & Sewell, 1998). More recently, Mittenberg, Patton, Canyock, & Condit (2002) reported widely varying rates (8–39%) depending upon the type of referral and whether or not the patient was in litigation. However, the former estimates were based on clinical judgment and the latter on widely varying criteria for determining cognitive symptom exaggeration, all of which suffer from the same methodological problems mentioned above.

Taxometric procedures, developed by Meehl (1995), have the potential to address these problems. These procedures were developed to identify whether a set of indicator scores results from the mixing of two latent classes or represent the observable result of a single underlying dimension. Taxometric analyses also provide a means for estimating the base rate of the latent category and the validity of indicators without pre-defined specification of category membership, something that previous studies were unable to do. These procedures use multiple consistency tests in order to identify a potential latent category (for an introduction see Cole, 2004; Lenzenweger, 2004).

2. The present study

The primary purpose of the present study was to examine whether VSVT indices measure a latent dimension or category of cognitive symptom exaggeration. This study used taxometric procedures to examine the underlying structure of the VSVT and to estimate the base rate of cognitive symptom exaggeration in typical outpatient neuropsychology settings. It was hypothesized that all VSVT measures would reveal an underlying taxonic structure. Taxonic structure was expected based upon the notion that individuals choose one of two approaches to cognitive tests, adequate effort/accurate presentation versus inadequate effort/negative response bias. In other words, our expectations of taxonic findings were based upon previous research suggesting that individuals who are feigning cognitive impairment adopt a strategy of invalid/inaccurate responding (Tan, Slick, Strauss, & Hultsch, 2002). It should be noted that this prediction of a decision-based taxon differs from the typical conceptualization (Haslam, 1997), but is consistent with conceptualizations regarding a possible psychopathological symptom exaggeration taxon (Strong et al., 2000). Furthermore, this conceptualization is supported by the relatively high rates of malingering observed in patients who are currently in litigation, suggesting that individuals may choose to exaggerate cognitive symptoms based upon the desire to gain external incentive or avoid criminal responsibility (Mittenberg et al., 2002). Thus, the present taxometric hypothesis, and resulting analyses, are driven by the rationale that the latent structure of cognitive symptom exaggeration is driven by patient decision-making.

Discerning the existence of a possible taxon, although academically interesting, also facilitates more accurate patient classification and base rate identification. The present study hypothesized that, if taxonic results were supported, the base rate of cognitive symptom exaggeration would be between 7 and 15%. This hypothesis is based upon the range of reported estimates for outpatient neuropsychological settings (Mittenberg et al., 2002; Rogers et al., 1993).

Researchers have noted that establishing the existence of a taxon is only the first step in the process, since identified taxa may have little meaning or represent pseudotaxa (Widiger, 2001). It is also important to establish the validity of taxa. Therefore, we generated the following expectations in the event that the taxonic hypothesis was supported.

First, taxon classifications were expected to be significantly correlated with performance on measures of intelligence and memory. Second, taxon members were expected to show greater increases in response latencies from easy to hard items relative to non-taxon, complement members. This prediction was derived from the notion that individuals exhibiting cognitive symptom exaggeration are consciously choosing the incorrect answers, a process that becomes more difficult as the choices become more similar. Third, taxon members were also expected to show higher levels of psychopathological symptom exaggeration on the Personality Assessment Inventory-Negative Impression Management scale relative to non-taxon members. This expectation is based upon previous work suggesting partial overlap of the cognitive and psychopathological symptom exaggeration constructs (Slick et al., 1996).

3. Method

3.1. Participants

Data for the present study were obtained from a de-identified patient registry that was reviewed and approved by the Institutional Review Board at the Cleveland Clinic Foundation. The database consisted of neuropsychological test data from adults referred for neuropsychological assessment at the Cleveland Clinic Foundation-Section of Neuropsychology. Approximately 50% of patient referrals were received from the Department of Neurology; approximately 25% from the Department of Psychiatry and Psychology; and approximately 25% from other sources, such as the Departments of Internal Medicine, Neurosurgery, Orthopedic Surgery, Rheumatology, Hematology/Oncology, Cardiology, and other departments or community sources. The most frequently occurring referral questions included assessment for possible dementia, attention-deficit/hyperactivity disorder, or learning difficulties; neurocognitive consequences of stroke, traumatic brain injury, or tumor; and pre-operative evaluations of patients considering epilepsy surgery. If participants were tested on more than one occasion, only data from an individual's initial evaluation were included in the present analyses. The sample consisted of 300 people (52% female; $M_{\text{age}} = 44.7$, $SD = 13.0$, range = 18–78). The racial distribution was consistent with that of patients seen in this setting (Caucasian 89%, African American 8%, Hispanic 1%, Asian 1%, Other 1%). On average, individuals had 13.8 years of education ($SD = 2.5$, range = 8–20), and 89.3% were right handed. None of the individuals included in the sample were seeking compensation at the time of the evaluation.

3.2. Measures

Individuals were administered a comprehensive battery of neuropsychological tests, including the VSVT. The VSVT is a computer administered 48-item, two-alternative, forced choice recognition memory measure designed to assess cognitive symptom exaggeration. For each item, individuals are shown a five-digit target stimulus followed by a delay. Then, two five-digit alternatives are presented, and the subject is instructed to select the response that matches the target stimulus. Items are presented in three blocks of 16 items. The three blocks represent increasing delays (5, 10, 15 s) between target and response presentation. Within each block, half of the response choices consist of foils with dramatic discrepancy from the target stimulus (easy trials) and half consist of foils with the transposition of only two digits (hard trials). Thus, the VSVT contains two manipulations, a delay manipulation and a difficulty manipulation. Response accuracy is recorded for the easy and hard trials of each delay period. Average response time (ms) and the standard deviation of response time (ms) are also recorded, separately for the 24 easy and 24 hard trials. However, preliminary analyses indicated that response time variables had substantial nuisance correlations. Therefore, this indicator set was not included in taxometric analyses. Reaction time measures were included in subsequent analyses to examine their potential clinical utility in the event the taxonic hypothesis was supported.

The Personality Assessment Inventory (PAI), a 344-item measure of personality and psychopathology was also administered to all individuals. Scores on the Negative Impression Management (NIM) were examined since this scale evaluates psychopathological symptom exaggeration. The Wechsler Adult Intelligence Scale (WAIS) and Wechsler Memory Scale (WMS) were also administered (Wechsler, 1997b). Full-scale IQ (FSIQ), verbal IQ (VIQ), performance IQ (PIQ), Auditory Immediate Memory Index (AIM), Auditory Delayed Memory Index (ADM), Auditory Delayed Recognition Memory Index (ADR), Visual Immediate Memory Index (VIM), Visual Delayed Memory Index (VDM), Immediate Memory Index (IM), General Memory Index (GM), and Working Memory Index (WM) were examined.

3.3. Analytic strategy

To examine the latent structure of the VSVT, three taxometric procedures were performed for each indicator set: mean above minus mean below a cut (MAMBAC), maximum eigenvalue at the hitmax point (MAXEIG), and latent mode (LMODE). Multiple taxometric procedures served as a consistency test for evaluating the taxonic conjecture. Taxometric analyses were performed using program code developed by Ruscio (2004b) for the R programming environment. Several decision rules were used to examine taxometric results: the nose count test based upon raters' evaluations of taxometric graphs (for an extended discussion of this test see Schmidt, Kotov, & Joiner, 2004), the variability of the base rate estimate identified in MAXEIG analyses, and the weighted fit d statistic provided in Ruscio (2004a) implementation of MAXEIG and MAMBAC. MAMBAC plots are concave when the data are generated from a dimensional latent structure, but convex when the data are taxonic. MAXEIG plots are irregular or flat when generated from dimensional structure, but convex when the data are taxonic. LMODE plots have a single large peak when the data are dimensional, but have two separate peaks when the data are taxonic. Three types of taxometric graphs were examined for each indicator set: the individual graphs, the average graph representing the average of all input indicator graphs, and the overlay graph in which the average graph is superimposed on the results from simulation data. For MAMBAC and MAXEIG, overlay graphs are presented because they contain the maximum amount of information. For LMODE, only a single graph is generated, with all of the input indicators contributing to this graph.

For the nose count test, three raters unfamiliar with taxometric procedures and blind to the present hypotheses rated the individual output indicator graphs resulting from each analysis, the average graph derived from the mean of all indicator graphs, and an overlay graph that presents the average graph superimposed on the average results from simulated data (± 1 SD). The overlay graph was derived from 20 simulated taxonic and 20 simulated dimensional data sets with roughly equivalent characteristics to the research data. For each taxometric procedure, raters were first presented with classic taxonic and dimensional graphs and given a verbal description of the general shape of taxonic and dimensional plots (i.e., MAMBAC taxonic plots are generally convex and shaped like a hill; and dimensional plots are generally concave and shaped like a valley). Raters were asked to place a 1 = taxonic, 2 = dimensional, or 3 = unspecified over each graph. Ratings were done separately for all MAMBAC graphs, followed by all MAXEIG graphs, and finally for all LMODE graphs. LMODE produced only individual and overlay graphs, because each indicator set generated only one individual graph, obviating the need for averaging.

The weighted fit d statistic examines the fit of research data to simulated dimensional and taxonic data. This index is highly similar to the comparison curve fit index, which has been shown to yield highly accurate discrimination of taxonic and dimensional data across a large range of data parameters, including sample sizes of 300 (Ruscio et al., in press). For the weighted d index, portions of the curves that best distinguish simulated taxonic from simulated dimensional data are given greater weight. Positive values indicate a relatively closer fit to simulated dimensional data and negative values indicate a relatively closer fit to simulated taxonic data (see Ruscio, 2004a). To compute these indices, 40 simulated data sets were generated for each analysis (20 dimensional, 20 taxonic) using the characteristics of the research data. This method has been previously used to evaluate the latent structure of psychopathic personality (Marcus, John, & Edens, 2004).

Convergence of base rate estimates should be observed when results have consistently identified a taxon. Although, there are no guidelines for interpreting the absolute magnitude of convergence, widely divergent results do not support the taxonic conjecture. Therefore, this decision rule was only applied in situations where other decision rules suggested possible taxonicity. To further examine the convergence of taxonic results, Bayesian classification of participants was requested for each MAXEIG analysis for which decision rules indicated taxonicity. These classifications were compared between indicator sets and to classifications derived from recommended VSVT cut scores (Slick et al., 1997).

3.4. Indicator sets

Three indicator sets were submitted to taxometric analyses. Multiple indicator sets present several advantages over the use of a single indicator set. First, they serve as an additional consistency test. Multiple indicator sets also decrease the possibility of missing potential taxa due to poorly specified or inappropriate indicators. In other words, the conclusion that a taxon is not present is weakly supported when only one indicator set is used, because the particular indicators employed may not have sufficient validity for taxon identification. In contrast, multiple indicator sets maximize the probability that the population of valid indicators for taxon/complement separation has been adequately

sampled. Previous research examining the taxonic conjecture has used multiple indicator sets (Ruscio, Ruscio, & Keane, 2002).

The three VSVT indicator sets examined were:

3.4.1. Total

The total number of correct responses for each of the delay periods was computed by summing the total number of correct responses to easy and hard items for each delay period. This yielded three indicators (total5, total10, total15); range of possible scores 0–16 for each indicator.

3.4.2. Hard

The total number of correct responses to the hard items for each of the delay periods (hard5, hard10, hard15) was examined; range of possible scores 0–8 for each indicator.

3.4.3. Easy

The total number of correct responses to the easy items for each of the delay periods (easy5, easy10, easy15) was examined; range of possible scores 0–8 for each indicator.

3.5. Suitability of indicator sets to taxometric procedures

There are several important considerations for determining whether a data set/indicator set is suitable for taxometric procedures. These include the overall sample size, the estimated taxon base rate, indicator skew, and indicator validity.

3.5.1. Overall sample size

Taxometric procedures tend to yield less interpretable results when the overall sample size is too low. Sample sizes of 300 have been recommended as providing clear results when other data parameters are adequate (Meehl, 1995), although other work has suggested that smaller sample sizes can be used for some taxometric procedures when other data parameters are good (Meehl & Yonce, 1994, 1996) and larger sample sizes may be needed when data parameters are poor (Schmidt et al., 2004). All of the indicator sets in this study had sample sizes of 300.

3.5.2. Taxon base rate

Having a moderate base rate is important for obtaining clear findings when using moderate sample sizes. The base rate of response distortion was conservatively estimated to be between .07 and .15 based upon previous work (Rogers et al., 1993). This modest base rate indicated that taxonic results may be missed, because results may have only indicated a half peak, rather than fully peaking and descending curves. Thus, the present analyses should be viewed as a conservative, and potentially under-powered, test of the taxonic hypothesis (Schmidt et al., 2004).

3.5.3. Indicator skew

Skewness has been shown to result in pseudotaxonicity (for an extended discussion of the effect of skewness on taxometric results see Schmidt et al., 2004). Therefore, before conducting taxometric analyses we examined the skew of all VSVT variables. Results indicated substantial negative skew for all accuracy variables (average skew = -3.55). To address this problem, accuracy variables were transformed using the LG10 function followed by inversion. All transformations were computed using SPSS (2002). Table 1 presents average skewness after transformation for each indicator set after. Even after transformation, the Easy indicator set still had substantial negative skew. Thus, results of

Table 1
Descriptive statistics and validity estimates for each indicator set

	Skew	Kurtosis	Sample r	Nuisance r	d
Total	−0.90	−0.71	0.61	0.30	1.56
Hard	−0.95	−0.67	0.63	0.26	2.80
Easy	−3.65	12.99	0.39	0.30	1.08

Note: These values are for transformed indicators.

the Easy indicator should be interpreted with caution. However, the other data sets had modest levels of skew following transformation. Analyses for all indicator sets were performed separately using raw and transformed data. The results were highly consistent. Therefore, for simplicity of presentation, and in the interest of presenting the most conservative test of the taxonic hypothesis, only results from the transformed data are presented.

3.5.4. Indicator validities

Some research has suggested that indicator validities of $d=1.2$ or greater produce reliable results and accurate classification. However, other studies have suggested that some taxometric procedures may yield interpretable results with indicator validities greater than $d=1.0$ (for a review see Schmidt et al., 2004). For this reason, we conducted a priori taxometric power analyses for each indicator set (Ruscio et al., 2002; Schmidt et al., 2004). These analyses computed estimates of average indicator validity using sample correlations, estimated nuisance correlations, and estimates of the taxon base rate. Nuisance correlations were estimated by computing the average correlation for the 24 individuals falling below the standard cutoff for the questionably valid range on the VSVT hard items and the average correlation for the upper 60% of the VSVT total accuracy or total latency distributions. The upper 60% was used to ensure that nuisance correlations were not unduly attenuated by range restriction. The base rate of negative response bias was expected to be below 40%, so no taxon members should have contributed to this estimate. Nuisance correlations were then averaged across indicators and sub-samples to provide a single nuisance correlation estimate for each indicator set. Estimates of average indicator validity were then computed using the formula provided in Meehl and Yonce (1996, p. 1146). These analyses are particularly useful for ruling out the possibility that negative findings result from low indicator validities of the taxon. Table 1 presents sample correlations, estimated nuisance correlations, average skewness, and d statistics for each indicator set. Inspection of Table 1 indicates that all indicator sets produced validities greater than $d=1.0$, and two of the three indicator sets had validities greater than 1.5. These results suggest that all indicator sets should produce clear results.

4. Results

4.1. Taxometric analyses

Table 2 presents taxometric decision rules, taxometric base rate estimates, and base rates derived from standard VSVT cutoffs. Fig. 1 presents overlay graphs for MAMBAC, MAXEIG, and LMODE analyses of all indicator sets.

4.1.1. MAMBAC

Results suggested taxonicity for all three indicator sets. Only two individual graph ratings to the Easy and one individual graph rating to the Total indicator set were unspecified and none were rated dimensional. All other ratings, including all average and overlay graph ratings, indicated taxonicity. The Total and Hard indicator sets produced highly similar base rates, but the Easy indicator set produced a much lower base rate. This was consistent with expectation, since very few individuals were expected to be susceptible to this aspect of the VSVT difficulty manipulation.

Table 2
Taxometric statistics and descriptives, separately by indicator set and taxometric procedure

Indicator	MAMBAC			MAXEIG				LMODE		VSVT	
	Rating	Base rate		Fit	Rating	Base rate		Fit	Rating	Base rate	
	% tax	<i>M</i>	<i>SD</i>		<i>d</i>	% tax	<i>M</i>		<i>SD</i>	<i>d</i>	% tax
Total	89	.22	.06	−4.16	100	.14	.01	−4.76	100	.21	.02
Hard	100	.26	.02	−2.68	100	.13	<.01	−4.22	100	.21	.08
Easy	78	.05	.05	−3.00	89	.02	.01	−4.81	100	.17	.00

Note: The VSVT base rate refers to the proportion of individuals who were classified as either invalid or questionably valid by standard cutoffs. Only ratings to individual graphs are presented.

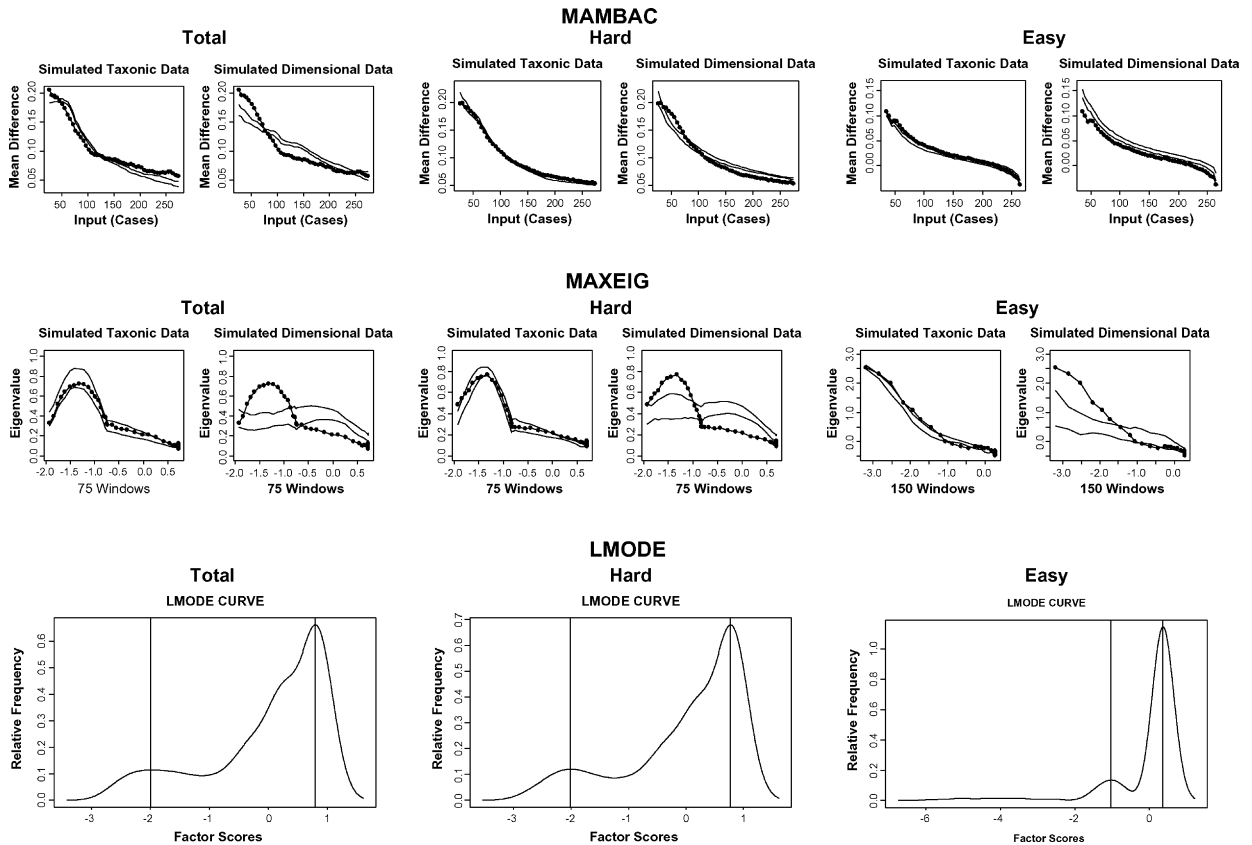


Fig. 1. MAMBA, MAXEIG, and LMODE overlay graphs for each indicator set.

4.1.2. MAXEIG

Results indicated clear evidence of taxonicity for the Total, Hard, and Easy indicator sets. Only one individual graph rating to the Easy indicator set was unspecified, none were dimensional, and all other ratings indicated taxonicity. The variability of base rates was quite small for all indicator sets, particularly for the Total and Hard indicator sets, and the weighted fit *d* statistic was highly negative, further supporting the taxonic conjecture. In general, MAXEIG results were highly consistent with those from MAMBA.

4.1.3. LMODE

Results indicated clear, second peaks for all indicator sets. Base rate estimates for the Total and Hard indicator sets were consistent with those from MAMBA and only slightly higher than those from MAXEIG.

4.1.4. Base rate comparisons

For the Total and Hard indicator sets, MAXEIG was the only procedure to show a clear full peak for each individual indicator graph. Therefore, it is likely that base rate estimates from MAXEIG (Total = .14 and Hard = .13) are more accurate than those from MAMBA or LMODE for these sets. MAMBA and MAXEIG showed good agreement for the Easy indicator set (.05 and .02), suggesting a very low base rate for this indicator set. Consistent with expectations, the base rates derived from MAXEIG and other taxometric procedures were higher than the base rates indicated by standard, binomial theory-based, VSVT classifications (see Table 2).

To examine the consistency of results across indicator sets, MAXEIG classifications for the Total, Hard, and Easy indicator sets were compared using Kappa coefficients and inspection of the 2 × 2 classification tables. MAXEIG classifications based upon the Total and Hard indicator sets showed good agreement ($\kappa = .87$). Agreement was lower for comparisons involving the Easy indicator set (largest $\kappa = .41$). This was due to the fact that the Easy indicator set identified fewer individuals as taxon members. However, no individuals identified as complement members by the

Total or Hard indicator sets were misidentified as taxon members by the Easy indicator set. Thus, the Easy indicator set appears to be suggesting a low base rate cognitive symptom exaggeration taxon composed of a sub-sample of individuals who miss both easy and hard items.

To compare the present results to existing classification methods, agreement between MAXEIG classifications, standard VSVT classifications, and those based upon a cut score of <90% to the hard items (Grote et al., 2000) was computed. For these comparisons, and for all future analyses examining the validity of taxometric classifications, MAXEIG classifications to the Hard indicator set were used since this indicator set showed the clearest peak, smallest base rate variability, and largest latent indicator validity. Results indicated that standard VSVT and <90% cut score classifications displayed modest agreement with taxometric classifications (VSVT $\kappa = .67$; <90% $\kappa = .75$). No individuals classified as invalid or questionably valid by standard VSVT classifications were deemed valid by MAXEIG. This finding is consistent with the lower base rate and insensitive nature of standard VSVT cutoffs (see Table 2). However, <90% cut score classifications produced a larger base rate of response distortion (22.3%). No individuals classified by MAXEIG as invalid were classified by <90% cut score as valid, although <90% cut score classified an additional 23 individuals as feigning. This is consistent with possible over-identification of response distortion using the <90% cut score (Grote et al., 2000; Haggerty, Frazier, Busch, & Naugle, 2006).

4.2. Taxon specification

To examine the hypotheses that taxon classifications would be significantly correlated with performance on the WAIS-III and WMS-III, point-by-serial correlations were computed for MAXEIG classifications to the Hard indicator set (see Table 3). A similar approach to potential taxon construct validation has been used in other recent studies (Kotov, Schmidt, Lerew, Joiner, & Ialongo, 2005). For these correlations, invalid performance (taxon group) was coded 1 and valid performance (complement group) was coded 2, with negative correlations expected. Results indicated that all correlations were significant (all p 's < .003). Group separation ranged from small to large ($d = -.32$ to $-.77$), with taxon members scoring 8–15 standard score on its lower on average than complement members.

To examine the hypothesis that taxon members are consciously choosing the incorrect answers, reaction times to easy and hard items were examined for both taxon and complement groups. We expected taxon members to show greater increases in reaction time from easy to hard items relative to complement members. Thus, we hypothesized an interaction between perceived item difficulty and taxon classification. To examine this hypothesis, a repeated measures analysis of variance was computed with taxon classification (taxon vs. complement) as the between subjects variable, item difficulty (easy vs. hard) as the repeated measures variable, and reaction time as the dependent variable. Results indicated significant main effects for taxon classification ($F(1, 298) = 136.1, p < .001$) and item difficulty ($F(1, 298) = 331.3, p < .001$). Taxon members yielded longer reaction times to both easy and hard items and hard items produced longer reaction times than easy items in both taxon and complement members. However, as predicted, these main effects were qualified by a significant interaction ($F(1, 298) = 91.3, p < .001$). This interaction indicated that taxon

Table 3
Correlations between MAXEIG and standard VSVT classifications and performance on the WAIS-III and WMS-III

		Taxon ($N = 44$)	Complement ($N = 256$)	MAXEIG	
		M (SD)	M (SD)	r	d
WAIS-III	VIQ	88.2 (15.2)	99.9 (14.4)	-.27	-.56
	PIQ	83.7 (14.7)	99.3 (13.9)	-.36	-.77
	FSIQ	85.2 (15.3)	99.6 (13.8)	-.34	-.72
WMS-III	AIM	88.2 (17.7)	98.9 (17.7)	-.23	-.47
	VIM	87.8 (17.5)	95.3 (16.2)	-.16	-.32
	IM	85.5 (19.6)	96.7 (16.6)	-.22	-.45
	ADM	88.4 (16.9)	99.7 (15.6)	-.24	-.49
	VDM	87.5 (17.6)	96.4 (16.4)	-.18	-.37
	ADR	87.0 (17.3)	100.6 (15.8)	-.28	-.58
	GM	85.4 (19.0)	98.4 (16.3)	-.26	-.54
	WM	88.4 (17.2)	98.9 (14.0)	-.24	-.49

Note: All correlations/effect sizes are significant, smallest $p < .003$.

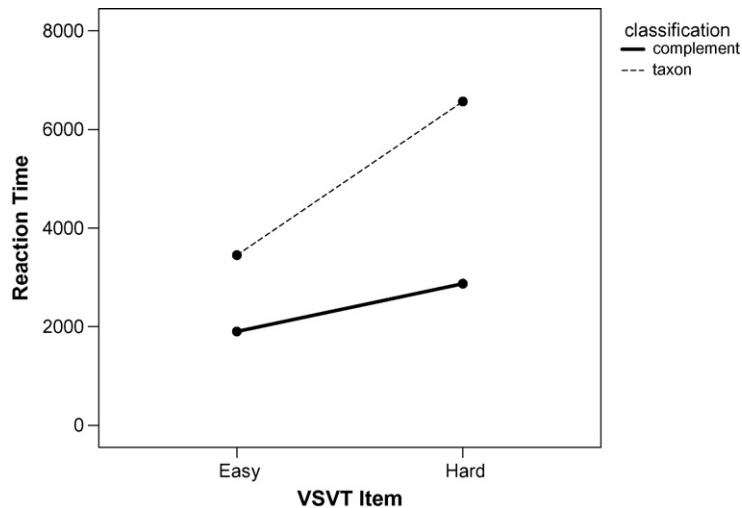


Fig. 2. Reaction times to easy and hard VSVT items, separately for taxon and complement members.

members showed a larger increase in reaction time from easy to hard items than complement members (see Fig. 2), consistent with the increased response time need to choose the incorrect answer to hard items.

To examine the relationship between cognitive and psychopathological symptom exaggeration, we computed the point-by-serial correlation between MAXEIG classifications and PAI negative impression management (NIM) scores. As expected, the correlation was significant, but modest in magnitude ($r = -.12$, $p = .021$), indicating taxon members showed higher NIM scores than complement members. Inspection of the data reveals that the taxon members with high scores on the PAI NIM tended to be individuals who scored poorly on both the Hard and Easy items of the VSVT. This indicates that some individuals show a very broad pattern of cognitive and psychopathological symptom exaggeration.

4.3. Clinical utility

Table 4 presents likelihood ratios, post-test probability, sensitivity, specificity, positive predictive power, and negative predictive power for VSVT hard accuracy and reaction time measures. MAXEIG classifications for the Hard indicator set were used as the criterion. Statistics are presented for only those areas of the accuracy total score distributions that showed overlap. For reaction time variables, statistics are presented for six percentile ranges, 1–30, 31–50, 51–70, 71–90, 91–95, and 96+, to increase clinical utility. Positive and negative predictive power was computed using the MAXEIG base rate estimate for the Hard indicator set (.13). Inspection of Table 4 reveals that hard accuracy total scores of 20 or 21 yield extremely low post-test probabilities of taxon membership, whereas scores of 18 or 19 are equivocal.

Table 4

Likelihood ratios, post-test probability, sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP) of VSVT accuracy and reaction time scores

	Likelihood ratio	Post-test probability	Sensitivity	Specificity	PPP	NPP
Hard accuracy						
18–19	7.88	.54	.99	.86	.38	.99
20–21	.32	.05	.95	.98	.78	.99
Hard reaction time						
<2297 ms	.04	.01	.98	.34	.18	.99
2298–2921 ms	.16	.02	.93	.57	.24	.98
2922–3498 ms	.26	.04	.86	.80	.39	.98
3495–5470 ms	3.12	.32	.48	.96	.64	.93
5470–7573 ms	10.71	.62	.27	.99	.80	.90
>7573 ms	31.60	.83	–	–	–	–

Reaction times below 3498 ms yield extremely low probabilities of taxon membership ($<.05$), whereas reaction times between 3495 and 7573 ms result in equivocal post-test classification, and scores greater than 7573 ms are associated with high, but not definitive, probabilities of taxon membership.

5. Discussion

This study identified a possible taxon defined by measures of cognitive symptom exaggeration in a sample of individuals referred for outpatient neuropsychological assessment. To ensure the findings were not accidental, multiple taxometric procedures were performed using multiple consistency tests and indicator sets. The taxonic conjecture passed all of the consistency tests with the clearest taxonic results converging on a base rate of .13–.14. However, it is worth noting that the present findings should only be viewed as tentatively supporting the taxonic hypothesis. This is because results of taxometric procedures, particularly those conducted in a single sample, should never be viewed as supporting the literal presence of a taxon (Watson, 2003). Future research using different, preferably larger, samples of patients followed longitudinally (to evaluate the stability of identified taxa) and using multiple, distinct measures of cognitive symptom exaggeration are needed to confirm the present findings.

With these caveats in mind, the present results have important implications for future research and clinical practice. First, support for the taxonic structure of VSVT accuracy variables permits estimation of the base rate of cognitive symptom exaggeration. The MAXEIG estimates of .13–.14 for the Total and Hard indicator sets are likely to be the most accurate estimates, since this estimate is based upon the clearest taxonic results. A base rate of .13–.14 is consistent with previous estimates (Mittenberg et al., 2002; Rogers et al., 1993) and indicates that negative response bias is a significant problem in outpatient settings. It should be noted, however, that base rate estimates will likely fluctuate depending upon the particular mix of patients presenting in the outpatient clinic. As the mix deviates from the mix in the present sample, the base rate will likely deviate as well. As noted by Mittenberg et al. (2002), settings with larger numbers of referrals for head injury may produce larger base rate estimates. The lower base rate estimates of .02–.05 for the Easy indicator set are likely due to the fact that these items are considerably less sensitive to cognitive symptom exaggeration, something that has been noted clinically. In fact, the major purpose of these items is to serve as a contrast to the purportedly hard items. Consistent with this notion, all of the individuals identified by this indicator set in the present study were also identified by the Hard indicator set. Therefore, findings for the Easy indicator set should be viewed as representing a sub-taxon of individuals who respond inaccurately to all item types and show the most blatant exaggeration.

Even with substantial variations in the base rate across settings, the present findings indicate that cognitive symptom exaggeration is a real problem. Unfortunately, many clinicians administer response bias measures sporadically or not at all. The present findings suggest that effort/symptom validity testing be used regularly, even in typical outpatient neuropsychological settings, since approximately 1 out of 7 non-litigating individuals presenting for evaluation will be exaggerating cognitive symptoms. However, it should be noted that the base rate estimates derived in the present study will likely under-represent the base rate of cognitive symptom exaggeration in forensic settings, where there is greater incentive to simulate cognitive problems. Additionally, the generalizability of the present findings will likely vary as a function of the similarity of the outpatient referral base and the type of symptom validity measure employed. Future research is needed employing taxometric procedures with different samples, particularly samples of individuals presenting in the context of litigation, to determine the magnitude of fluctuations in the base rate of cognitive symptom exaggeration. This research should also examine different types of symptom validity measures to better assess the generalizability of the present findings.

Second, if the present results are confirmed in future studies, the taxonic structure of VSVT accuracy measures permits more accurate classification of individuals. Currently, the VSVT classifies individuals into three score ranges: valid (total raw score 30–48, easy and hard raw scores 16–24), questionably valid (total raw score 18–29, easy and hard raw scores 8–15), and invalid (total raw score 0–17, easy and hard raw score 0–7). In the present sample, only three individuals fell in the invalid range ($<1\%$) on hard items. Combining the invalid and questionably valid ranges increased the percentage of individuals detected (8%), but numerous individuals were still missed. Thus, the hypothesis that current VSVT score ranges are too conservative received strong support. This is likely due to a desire on the part of the test developers to maximize positive predictive power as well as heavy reliance on binomial probability theory for identifying individuals who produce patterns that are clearly inconsistent with adequate effort. Similarly, $<90\%$ cut score classifications resulted in a large number of individuals classified as feigning (67 of 300), resulting in an unlikely

base rate of response bias for this outpatient setting. This finding, coupled with previous research (Grote et al., 2000), suggests that this cut score is too liberal and its use clinically would result in substantial amounts of valid cognitive test data going uninterrupted.

Future revisions of cognitive symptom validity measures may benefit from the adjunctive use of Bayesian estimation based upon the results of taxometric procedures, such as the estimates presented in this study. This approach is likely to classify individuals more accurately by utilizing all of the information in the data set to make classification decisions. For clinicians familiar with the use of likelihood ratios and for whom good estimates of the base rate of cognitive symptom exaggeration are available, the information presented in Table 4 can be used to compute the probability that an individual is exerting inadequate effort. However, it should be noted that these data may be less useful in settings with very different referral patterns. If future research conducted in different settings supports the findings of the present study, it is likely that Bayesian classification based upon taxometric results could be used in legal settings in a similar fashion to medical test results. In particular, forensic neuropsychologists could benefit from this approach, since it presents an easily understood interpretation of test performance. For example, an individual with a hard accuracy score of 16 out of 24 could be characterized as >99% likely to have exhibited inadequate effort on the VSVT. Unfortunately, using binomial probability theory as the guideline, this individual would have fallen at the low end of the valid range, further demonstrating the insensitivity of this criterion. Alternatively, an individual scoring 20 out of 24 would be classified as unlikely to be exaggerating cognitive symptoms, while the >90% cut score would place this individual in the invalid category.

Lastly, the present findings clarify the construct validity of cognitive symptom exaggeration as a particular kind of response set, qualitatively different from adequate effort. This is in contrast to the common clinical interpretation of symptom validity tests as being indicators of a continuum of effort. To further clarify this qualitative distinction, taxon members showed substantially lower scores (8–15 points lower) on all memory and IQ indices than complement members. Interestingly, the largest separation was observed for IQ measures indicating that the VSVT is not just sensitive to memory dissimulation, but to a broader array of exaggeration.

Taxon members also showed larger increases in response latencies from easy to hard items than complement members. This finding is consistent with the idea that taxon members are consciously processing both the correct and incorrect responses and then choosing the incorrect response. In other words, taxon members may show particularly long reaction times to hard VSVT items because response choices for hard items differ by only the transposition of two digits, requiring longer processing time to choose the incorrect answer. Complement members, on the other hand, only need to identify the number they learned and do not need to distinguish between the two numbers. Future work is needed to replicate this finding and more carefully examine the cognitive processes involved in cognitive symptom exaggeration.

Taxon members also tended to show higher levels of negative impression management. This is consistent with previous research indicating that some individuals showing cognitive symptom exaggeration also show psychopathological symptom exaggeration (Slick et al., 1996). Future work is needed to more accurately characterize the patterns of cognitive or affective symptoms endorsed by these individuals. It may be that individuals who show response bias on both measures are reporting more generalized cognitive and emotional dysfunction.

The present findings also yield important information for clinicians. Specifically, VSVT accuracy and latency raw scores provide good separation of individuals who are feigning cognitive symptoms from those who are not. For VSVT hard accuracy, taxon and complement groups only showed overlap at four points of the raw score distribution (18–21) and likelihood ratios were still quite useful for ruling out cognitive symptom exaggeration in the upper portion of this range. Future research is needed with other clinically referred groups of litigants and non-litigants to determine whether the lack of overlap between taxon and complement members is consistent across samples. It may be that overlap will increase (decreases in group separation) as individuals are educated about symptom validity testing, necessitating the development of more refined measures or the use of multiple measures in an evidence-based framework (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000).

Future studies should use multiple measures of cognitive and psychopathological symptom exaggeration to clarify the convergent and discriminant validity of these constructs. This will assist in determining the level of overlap between these two broad constructs and identifying the characteristics of individuals who show poor performance on multiple or only one type of negative response bias measure. For example, some individuals may show high scores on all measures of psychopathological and cognitive symptom exaggeration, whereas others may show a more nuanced pattern. Future research will also be helpful to more accurately characterize the specific types or dimensions of

response bias underlying the broader constructs. It is likely that the VSVT only assesses some aspects of cognitive symptom exaggeration. Inclusion of multiple measures of cognitive symptom exaggeration will permit more accurate identification of the dimensions or types of response bias observed in typical outpatient assessment settings. This work should examine varied measures of response bias (computer vs. person administered, visual vs. verbal presentation, etc.) and account for method variance in a systematic fashion.

Distinct facets of cognitive symptom exaggeration may also yield incremental validity over more general measures. Specifically, combining multiple facets or measures of cognitive symptom exaggeration will likely improve the accuracy of classification of possible response bias taxa. Specification of distinct categories or dimensions of cognitive symptom exaggeration will also help identify the particular cognitive domains affected by certain patterns of response bias. At present, poor scores on measures of cognitive response bias are typically assumed to indicate poor performance on all other cognitive measures, but this may not be true. Instead, different aspects of cognitive symptom exaggeration may have differential predictive validity. For example, accuracy scores may show stronger relationships with performance on declarative memory tasks, such as the information subtest of the WAIS-III (Wechsler, 1997a), whereas latency scores may show stronger relationships with speeded tasks, such as digit symbol. Similarly, individuals performing poorly on verbal, memory-based measures of symptom exaggeration may produce different patterns of neuropsychological deficits than individuals performing poorly on visual, memory-based measures of response bias. Future studies are needed to examine these possible distinctions so that they can be accounted for in neuropsychological test interpretation. These studies will also help clinicians to select appropriate response bias measures, so that all possible distortions to the test data can be quantified.

5.1. Limitations and future directions

There are a few limitations to the present study. First, the sample size was modest and at the lower end of the recommended level. Moderate sample size, coupled with the small to moderate base rate of the taxon, likely contributed to over-estimation of the base rate in MAMBAC. This is because the end slabs (the first point estimates in taxometric graphs) may have included some complement members, causing incomplete peaking of the graphs. Future studies should examine populations that are likely to have higher base rates of cognitive symptom exaggeration, such as forensic neuropsychological settings, in order to avoid this problem. However, the fact that clear taxonomic findings were observed for all indicator sets suggests that indicator validity is quite high and modest sample sizes may be adequate for examining these types of indicators.

References

- Arbisi, P. A., & Ben-Porath, Y. S. (1998). The ability of the Minnesota Multiphasic Personality Inventory-2 validity scales to detect fake-bad responses in psychiatric inpatients. *Psychological Assessment, 10*(3), 221–228.
- Bender, S. D., & Rogers, R. (2004). Detection of neurocognitive feigning: Development of a multi-strategy assessment. *Archives of Clinical Neuropsychology, 19*(1), 49–60.
- Berry, D. T. R., & Butcher, J. N. (1998). Detection of feigning of head injury symptoms on the MMPI-2. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation. Critical issues in neuropsychology* (pp. 209–238 xii, 291).
- Cole, D. A. (2004). Taxometrics in psychopathology research: An introduction to some of the procedures and related methodological issues. *Journal of Abnormal Psychology, 113*(1), 3–9.
- Doss, R. C., Chelune, G. J., & Naugle, R. I. (1999). Victoria Symptom Validity Test: Compensation-seeking vs. non-compensation-seeking patients in a general clinical setting. *Journal of Forensic Neuropsychology, 1*(4), 5–20.
- Franzen, M. D., & Iverson, G. L. (1998). Detecting negative response bias and diagnosing malingering: The dissimulation exam. In P. J. Snyder & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment*. Washington, DC: American Psychological Association.
- Frederick, R. I. (2002). Review of the validity indicator profile. *Journal of Forensic Neuropsychology, 2*, 125–145.
- Greene, R. L. (1997). Assessment of malingering and defensiveness by multiscale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 169–207).
- Grote, C. L., Kooker, E. K., Garron, D. C., 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 & Experimental Neuropsychology, 22*(6), 709–719.
- Haggerty, K. A., Frazier, T. W., Busch, R. M., & Naugle, R. I. (submitted for publication). *Relationships Between Victoria Symptom Validity Test Indices and Personality Assessment Inventory Validity Scales in a Large Clinical Sample*.
- Haslam, N. (1997). Evidence that male sexual orientation is a matter of degree. *Journal of Personality and Social Psychology, 73*, 862–870.

- Kotov, R., Schmidt, N. B., Lerew, D. R., Joiner, T. E., & Ialongo, N. S. (2005). Latent structure of anxiety: Taxometric exploration. *Psychological Assessment, 17*(3), 369–374.
- Langeluddecke, P. M., & Lucas, S. K. (2003). Quantitative measures of memory malingering on the Wechsler Memory Scale-Third Edition in mild head injury litigants. *Archives of Clinical Neuropsychology, 18*(2), 181–197.
- Langeluddecke, P. M., & Lucas, S. K. (2004). Validation of the rarely missed index (RMI) in detecting memory malingering in mild head injury litigants. *Journal of Forensic Neuropsychology, 4*(1), 49–64.
- Lenzenweger, M. F. (2004). Consideration of the challenges, complications, and pitfalls of taxometric analysis. *Journal of Abnormal Psychology, 113*(1), 10–23.
- Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment, 14*(2), 170–176.
- Marcus, D. K., John, S. L., & Edens, J. F. (2004). A taxometric analysis of psychopathic personality. *Journal of Abnormal Psychology, 113*(4), 626–635.
- Meehl, P. E. (1995). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–275.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194–216.
- Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74*, 1059–1274.
- Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*, 1091–1227.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical & Experimental Neuropsychology, 24*(8), 1094–1102.
- Morey, L. C. (1991). *Personality assessment inventory*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (2003). *Essentials of PAI assessment*. New York, NY: John Wiley & Sons, Inc..
- Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9*(6), 501–552.
- Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13*, 255–274.
- Rogers, R., Salekin, R. T., & Sewell, K. W. (1998). A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior, 22*(4), 353–367.
- Rogers, R., Sewell, K. W., Cruise, K. R., Wang, E. W., & Ustad, K. L. (1998). The PAI and feigning: A cautionary note on its use in forensic-correctional settings. *Assessment, 5*(4), 399–405.
- Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*(4), 349–359.
- Ruscio, A. M., Ruscio, J., & Keane, T. M. (2002). The latent structure of Posttraumatic Stress Disorder: A taxometric investigation of reactions to extreme stress. *Journal of Abnormal Psychology, 111*(2), 290–301.
- Ruscio, J. (2004a). Taxometric Program Documentation, R language. Retrieved December 30, 2004, from <http://www.etown.edu/psychology/Faculty/Ruscio.htm>.
- Ruscio, J. (2004b). Taxometric Programs in the R language. Retrieved December 30, 2004, from <http://www.etown.edu/psychology/Faculty/Ruscio.htm>.
- Ruscio, J., Ruscio, A. M., & Meron, M. (in press). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research*.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). Edinburgh: Churchill Livingstone.
- Schmidt, N. B., Kotov, R., & Joiner, T. E. (2004). *Taxometrics: Toward a new diagnostic scheme for psychopathology*. Washington, DC: American Psychological Association.
- Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. J. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical & Experimental Neuropsychology, 18*(6), 911–922.
- Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *VSVT: Victoria Symptom Validity Test, Version 1.0. Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc..
- Slick, D., 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*(4), 545–561.
- Slick, D. J., Tan, J. E., Strauss, E., Mateer, C. A., Harnadek, M., & Sherman, E. M. S. (2003). Victoria Symptom Validity Test scores of patients with profound memory impairment: Non-litigant case studies. *Clinical Neuropsychologist, 17*(3), 390–394.
- SPSS. (2002). *SPSS Professional Statistical Package (Version 11.0)*. Chicago, Illinois: SPSS Inc..
- Storm, J., & Graham, J. R. (2000). Detection of coached general malingering on the MMPI-2. *Psychological Assessment, 12*(2), 158–165.
- Strong, D. R., Greene, R. L., & Schinka, J. A. (2000). A taxometric analysis of MMPI-2 infrequency scales [F and F(p)] in clinical settings. *Psychological Assessment, 12*(2), 166–173.
- Tan, J. E., Slick, D., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *Clinical Neuropsychologist, 16*(4), 495–505.
- Thompson, G. B. (2002). The Victoria Symptom Validity Test: An enhanced test of symptom validity. *Special Issue: Detection of response bias in forensic neuropsychology: Part I, 2*(3–4), 43–67.

- Watson, D. (2003). Investigating the construct validity of the dissociative taxon: Stability analyses and normal and pathological dissociation. *Journal of Abnormal Psychology, 112*(2), 298–305.
- Webb, L. M. (1999). Clinical assessment of malingering utilizing the Minnesota Multiphasic Personality Inventory-II (MMPI-II), Millon Clinical Multiaxial Inventory-III (MCMI-III), and Dissociative Experiences Scale.
- Wechsler, D. (1997a). *Manual for the Wechsler Adult Intelligence Scale-Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997b). *WAIS-III WMS-III technical manual*. San Antonio, TX: The Psychological Corporation.
- Widiger, T. A. (2001). What can be learned from taxometric analyses? *Clinical Psychology: Science and Practice, 8*, 528–533.