2014

Multimethod Assessment of Distortion: Integrating Data from Interviews, Collateral Records, and Standardized Assessment Tools

Danielle Burchett
California State University, Monterey Bay, dburchett@csumb.edu

R. Michael Bagby

Recommended Citation

Burchett, D., & Bagby, R. M. (2014). Multimethod assessment of response distortion: Integrating data from interviews, collateral records, and standardized assessment tools. In C. Hopwood & R. Bornstein (Eds.), Multimethod clinical assessment (pp. 345-378). New York, NY: Guilford.

Multimethod Assessment of Distortion

Integrating Data from Interviews, Collateral Records, and Standardized Assessment Tools

Danielle Burchett and R. Michael Bagby

Psychological assessment is an endeavor that has the potential to provide critical information to the referral source or other users of such service, including clinicians from other professions (e.g., psychiatrists, social workers, nurses), teachers, employers, and those in the legal arena (e.g., lawyers, judges, correctional officers). Yet, a variety of factors may undermine the validity and usefulness of the outcomes of psychological assessments. There are, for example, the inherent psychometric limitations of the instruments used in an assessment battery, and in many assessment contexts, the utility of psychological test results may be compromised by test response bias (e.g., overreporting or underreporting). For instance, scores from scales designed to assess different forms of psychopathology may be artificially elevated or deflated and, therefore, less predictive of the constructs they were meant to measure due to non–content-based or content-based invalid responding (Burchett & Ben-Porath, 2010; Wiggins, Wygant, Hoelzle, & Gervais, 2012). Thus, when conducting personality assessments, it is critical to examine validity scales—indices of response distortion—to determine whether inaccurate symptom reporting influenced the accuracy of the scales of greatest clinical interest: the substantive measures of psychopathology and personality dysfunction.

In addition to providing information about the predictive validity of substantive scale test scores, validity scales can also provide clinically relevant data about the accuracy of examinee self-report. For instance, overreporting in a clinical setting may be indicative of a "cry for help" (Graham,

2006), a stable personality style (Morey, 2007), or intentional distortion for secondary gain (American Psychiatric Association, 2013). Thus, the assessment of response distortion has two main functions: (1) to determine whether substantive test results are valid measures of genuine symptoms, and (2) to provide information about an examinee's presentation that could impact the course of treatment or legal decisions.

In this chapter we first review different types of invalid responding, symptom domains in which they occur, and extant models explaining their etiology. Next, we address strategies and tools used for the detection of misleading responding and then conclude by recommending multimethod assessment of response distortion and providing an illustrative case example.

Types of Response Distortion

Ben-Porath (2003; Table 12.1) discussed the threats to protocol validity in self-report personality assessment, including *non–content-based* and *content-based* invalid responding. Non–content-based invalid responding

TABLE 12.1. Response Styles That May Invalidate Personality and Psychopathology Assessment Results

Non–content-based invalid (NCBI) responding	
Intentional NCBI responding	Unintentional NCBI responding
Intentional nonresponding	Unintentional nonresponding
Intentional random responding	Unintentional random responding
Intentional fixed responding	Unintentional fixed responding
Intentional acquiescence	Unintentional acquiescence
Intentional counteracquiescence	Unintentional counteracquiescence

Content-based invalid responding	
Intentional overreporting	Unintentional overreporting
Feigning (Rogers & Bender, 2003)	Poor insight
Exaggeration	Negative emotionality
Fabrication	
Malingering (American	
Psychiatric Association, 2013)	
Intentional underreporting	Unintentional underreporting
Impression management	Social desirability/self-deception
Minimization	
Denial	
Defensiveness (Rogers, 1984)	

Dissimulation (Rogers, 2008a)

Hybrid responding (Rogers, 2008a)

Note. Expanded from Ben-Porath (2003).

occurs when an individual's invalid test responses are unrelated to the item content. *Nonresponding, random responding,* and *fixed responding* also compromise protocol validity. Nonresponding occurs when the examinee does not respond to test items, whereas random responding occurs when an examinee indiscriminately responds to items with varying answers (e.g., a random true–false–false–true–true pattern). Fixed responding involves indiscriminately providing the same response to test items and includes *acquiescence* (e.g., all "true" responding) and *counteracquiescence* (e.g., all "false" responding). These response styles occur on Likert-type scales as well, with examinees randomly responding or choosing responses at the same level without considering item content (Ben-Porath, 2003).

When an examinee pays attention to the content of the items but provides responses that depict a distorted picture of their actual functioning, it is commonly referred to as *content-based invalid responding.* This may be intentional or unintentional and includes *overreporting* and *underreporting* of symptoms.[1] *Overreporting* occurs when an examinee's responses lead them to appear worse off than they actually are. *Feigning* (Rogers & Bender, 2003) is the intentional exaggeration or fabrication of symptoms (making no assumption as to the examinee's motivation). *Malingering* is a subcategory of feigning, which is by definition externally motivated, context-specific intentional overreporting of symptoms (American Psychiatric Association, 2013). Unintentional overreporting may occur due to *poor insight* into one's symptoms or *negative emotionality,* which predisposes individuals to believe they are more impaired than is true for them (Ben-Porath, 2003; Tellegen, 1985).

Underreporting occurs when an individual's response style leads them to appear better off than they actually are. *Impression management* involves the intentional minimization (reporting fewer symptoms than are present and/or minimizing the severity of those that are reported) or *denial* (denying all symptoms) of problems (Ben-Porath, 2003) in an attempt to create a positive image or demonstrate mental health. Rogers (1984) described *defensiveness* as the opposite of malingering: intentional minimization of symptoms. Unintentional underreporting has been labeled *self-deception* and *social desirability* and is believed to be an inadvertent masking of symptoms (Strong, Greene, Hoppe, Johnston, & Olesen, 1999). Intentional impression management is considered to be a setting-specific strategy, whereas self-deception is thought to be a stable personality trait

---

[1] Rogers, Sewell, and Gillard (2010) recommend against the labels "overreporting" and "underreporting" because they are nonspecific. They instead recommend "dissimulation" to describe intentional response distortion and "feigning" to describe intentional symptom exaggeration. We retain the terms *overreporting* and *underreporting* in this chapter precisely because test results cannot speak to intent. We recommend that the evaluator specify intent and symptom type (e.g., cognitive, somatic, psychopathology) when each can be determined, and we caution that intent cannot be determined by the test results but rather by other collateral information.

(Paulhus, 1988). Rogers (2008a) notes that it is possible to observe *hybrid responding*, in which an examinee utilizes more than one response style (e.g., honest responding in most domains but underreporting of substance abuse).

Domains of Response Distortion

Content-based invalid responding may occur in three general symptom domains: reported somatic complaints, cognitive complaints, and psychopathology (e.g., Hoelzle, Nelson, & Arbisi, 2012; Rogers, Sewell, & Gillard, 2010). Although the detection of somatic and cognitive response distortion is important, it is beyond the scope of this book.[2] The third domain in which invalid responding may occur—in the reporting of psychopathology—is the focus of this chapter. For brevity, we did not include a discussion of performance-based personality tools (e.g., Rorschach and Thematic Apperception Test).

Models of Response Distortion

Several models have been proposed to explain the source of invalid responding. These models generally fall into three main categories: those that propose (1) underlying psychopathology (e.g., pathogenic model, Bash & Alpert, 1980; interpersonal management model, Heinze, 1999); (2) psychopathy, antisocial personality, and criminal behavior (e.g., criminological model; American Psychiatric Association, 1980); or (3) cost–benefit analysis of potential risks and gains in conjunction with individual factors (e.g., adaptational model, Rogers & Cavanaugh, 1983; interactional model of applicant faking, Snell, Sydell, & Lueke, 1999; interactionist model of item-level response distortion, Tett et al., 2006) as reasons for response distortion (see Rogers, 1990, 2008a).

Although there has been support for the adaptational model (e.g., Thomas-Peter, Jones, Campbell, & Oliver, 2000), DSM-5 (American Psychiatric Association, 2013) subscribes to the criminological model. It defines malingering as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (p. 726), and ignores underreporting altogether. According to the DSM, any combination of the following should arouse suspicion of malingering: (1) medicolegal context of presentation, (2) marked discrepancy between claimed stress or disability and objective findings, (3) lack

---

[2] The interested reader is referred to several sources that review the assessment methods within these domains (e.g., Boone, 2007; Hall & Poirier, 2001; Larrabee, 2007; Morgan & Sweet, 2009).

of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen, or (4) presence of antisocial personality disorder. Research indicates that these markers are sensitive, but not specific, predictors of malingering (e.g., Kucharksi, Duncan, Egan, & Falkenbach, 2006; Vitacco, 2008).

Berry and Nelson (2010) outlined additional concerns with the (then) DSM-IV-TR guidelines. For instance, the DSM does not distinguish between domains of overreporting (somatic, cognitive, psychopathology), each of which should be assessed using different techniques. Additionally, the DSM calls for the examiner to determine *intent* and *motivation*, which cannot be assessed by psychopathology overreporting tests and can only sometimes be determined with self-report or collateral information (Rogers et al., 2010). Further, Berry and Nelson (2010) discuss the difficulty in distinguishing between *externally motivated* malingering and *internally motivated* somatoform disorder.

Berry and Nelson (2010) recommended changes to the DSM Malingering V code. Within the psychopathology domain, they recommended a focus on (1) using well-validated techniques without attempting to infer intent or motivation, (2) considering literature on the utility of various detection strategies, (3) employing multiple detection strategies to increase accuracy and minimize false-positive rates, (4) specifying the domain(s) in which false symptoms were documented, and (5) documenting the evaluator's level of certainty (possible, probable, definite) and the severity (mild, moderate, severe) of overreporting. We echo these suggestions and believe clinicians should utilize a variety of well-validated detection strategies in addition to collateral information and behavioral observations to assess for overreporting *and* underreporting across the three domains.

Strategies for Detecting Response Distortion in Personality Assessment

Because examinees may distort their clinical presentation in a variety of ways, several detection strategies have been developed. It is important to understand differences between these strategies because useful multimethod assessment of response distortion will employ several strategies during any given assessment.

Detecting Overreported Psychopathology

Numerous methods have been used to develop validity scales and standalone instruments to detect overreported psychiatric symptoms such as psychosis, depression, or anxiety. *Quasi-rare symptoms* scales are made of items reflecting symptoms rarely endorsed in normative samples. Although such scales differentiate between "normal" individuals and those who

report severe problems, it can be unclear whether elevations are due to overreporting or genuine psychopathology. *Rare symptoms* scales address this concern by including items reflecting symptoms rarely endorsed in clinical samples. These types of scales are much less confounded by genuine psychopathology. The *improbable symptoms* approach is similar to the rare symptoms approach, except that items are of preposterous or ridiculous nature. Clearly, individuals endorsing many of these items are inaccurately reporting their symptomatology. However, these items tend to be so improbable that even moderately sophisticated malingerers can detect and avoid endorsing them (Rogers, 2008b).

The *symptom combinations* strategy involves creating items involving two symptoms, which may commonly occur but rarely occur in combination. The *spurious patterns of psychopathology* method is similar, but includes item combinations reported by malingerers but not genuine patients. The *indiscriminant symptom endorsement* approach was created with the assumption that malingerers tend to endorse an overall higher rate of symptoms than do individuals with genuine psychopathology. *Symptom severity* scales measure the severity or number of severe symptoms endorsed. Those who endorse many of these items as severe are likely to be overreporting symptoms. The *obvious symptoms* detection method uses face-valid symptoms either alone or in combination with more subtle symptoms to differentiate between genuine patients and overreporters (who are more likely to endorse many symptoms that are obviously related to psychopathology). The *reported versus observed* method measures differences between self-reported problems and clinical observations, with the assumption that individuals who report multiple problems not seen by clinicians are exaggerating or fabricating symptoms. The *erroneous stereotypes* detection method involves asking about symptoms that individuals often believe are related to true psychopathology, but in fact, are not (Rogers, 2008b).

Detecting Underreported Psychopathology

A number of detection strategies have been created to determine when an examinee is minimizing problems or exaggerating positive qualities (Rogers, 2008b). Some validity scales utilize a *denial of minor flaws/personal faults* strategy to detect underreporting. Individuals who do not admit to such minor flaws are likely intentionally presenting themselves in a favorable light (Graham, 2006). The *spurious patterns of simulated adjustment* strategy identifies scale configurations that occur in defensive patients but that are uncommon in clinical and community samples. The *denial of psychopathology/patient characteristics* method uses items that differentiate between individuals with known psychopathology who score within normal limits (and are thus believed to be underreporting) and normative samples. The goal of this method is to distinguish between those who *are*

generally asymptomatic and those who have symptoms but deny them. The *social desirability* method attempts to identify examinees who present with a highly favorable image. Rogers (2008b) also described a *blended affirmation of virtuous behavior and denial of personal faults* method that combines items about virtuous behaviors and personal faults on the same scale.

Use of Multiple Methods to Detect Distortion

Given the complex nature of response distortion, many researchers have advocated for a multifaceted approach to its detection (e.g., Bender & Rogers, 2004; Mihura, 2012; Ray, 2009). Below we discuss the information that can be gained from clinical interviews, behavioral observations, and collateral sources and provide information about several tools used to detect response distortion. These scales and the domains of response distortion they purport to measure are displayed in Table 12.2.

Clinical Interview and Behavioral Observations

Ironically, perhaps the best-known advocate for the utility of standardized assessment methods, Paul Meehl (1996) stated, "if I were asked to diagnose a mental patient and told that I could either have an MMPI profile or conduct a mental status examination, I would prefer the latter." He appreciated the immensely rich information that can be gathered from a discussion with the examinee and from observing their behaviors. In the context of response distortion, it is important to be attuned to patterns of behavior inconsistent with major mental health conditions and observe whether distractibility, personality patterns, or other factors may be contributing to inaccurate test results. Further, when the setting permits, it can be helpful to observe the examinee's behaviors outside of the formal testing session because they may not think to distort their presentation at those times. Observing discrepancies in and out of the testing setting may provide evidence that the examinee's distortion is *intentional*.

Collateral Information

Treatment records and discussions with staff and family members may provide important background information about an examinee's history. Records may document mental status and behaviors around the time of the evaluation and may help inform whether behaviors are inconsistent with self-reported problems during the evaluation. Of note, while family members may provide information about the examinee's mental health history, they may have agendas in support of (or against) their relative or a poor understanding of the examinee's mental health history. Finally,

TABLE 12.2. Domains of Response Distortion Measured by Common Personality and Psychopathology Inventories and Standalone Measures

	Non-content-based invalid responding			Overreporting			Underreporting	
	Nonresponding	Random responding	Fixed responding	Psychopathology	Cognitive complaints	Somatic complaints	Uncommon virtues	Denial of psychopathology
MMPI-2	CNS/(?)	VRIN	TRIN	F, FB, FP	FBS	FBS	L, S	K, S
MMPI-2-RF	CNS/(?)	VRIN-r	TRIN-r	F-r, FP-r	FBS-r, RBS	FBS-r, FS	L-r	K-r
PAI	Missing items	ICN, INF		NIM, MAL, RDF		MPRDF	PIM, DEF, CDF	DEF, CDF
MCMI-III		V		X, Z			X (low), Y	X (low), Y
SNAP		VRIN, II	TRIN, II	DEV, II			DRIN, RV, II	DRIN, II
SNAP-2		VRIN, II	TRIN, II	DEV, BDEV, II	IF		DRIN, RV, II	DRIN, II
NEO-PI-R	Response count; items A, B, and C	Response Count; items A, B, and C	Response count; items A, B, and C	Item A			Item A	Item A
SIRS		INC		RS, SC, IA, BL, SU, SEL, SEV, RO, DA, OS, SO			DS	
SIRS-2		INC		RS, SC, IA, BL, SU, SEL, SEV, RO, DA, OS, RS, Total, MT Index, SS Index			DS	
SIMS				P, Af				N, Am, LI
M-FAST				RO, ES, RC, UH, USC, NI, S, Total Score				

Note. Refer to the text for full-scale names and descriptions. II is a composite score indicating protocol invalidity and is elevated in part due to several types of response distortion.

understanding contextual information (e.g., whether evaluation results will impact the examinee's freedom), it is possible to identify *potentially* motivating factors.

Embedded Validity Indices on Self-Report Measures

Minnesota Multiphasic Personality Inventory–2[3]

The Minnesota Multiphasic Personality Inventory–2 (MMPI-2; Butcher et al., 1989, 2001) is a 567-item self-report personality and psychopathology inventory. It includes three scales designed to measure non–content-based invalid responding. Cannot Say (CNS/?) is a count of the number of items omitted or marked as *both* true and false. If many items are omitted or double-marked, the validity of the scale scores is called into question because substantive scale scores will be based on incomplete information. Variable Response Inconsistency (VRIN), designed to measure variable responding, consists of item pairs with similar or opposite content. Points are scored when individuals endorse these pairs in an empirically and conceptually inconsistent manner. True Response Inconsistency (TRIN) consists of item pairs opposite in content. Points are scored when individuals answer both items in the same direction. It is designed to measure both acquiescence and counteracquiescence.

The MMPI-2 also includes several overreporting indices. Infrequency (F), developed using the *quasi-rare symptoms* strategy, consists of items rarely endorsed in the MMPI-2 normative sample. Also developed using the *quasi-rare symptoms* approach, Back Infrequency (F_B) supplements F because it consists of infrequently endorsed items found in the latter part of the MMPI-2 booklet. Infrequency Psychopathology (F_P; Arbisi & Ben-Porath, 1995), developed using the *rare symptoms* approach, consists of items infrequently endorsed in both the MMPI-2 normative sample and by psychiatric inpatients. Its design makes it less likely than F to be elevated due to psychopathology, resulting in fewer false-positive results. Symptom Validity (FBS; Lees-Haley, English, & Glenn, 1991) was developed from rational item selection using the *unusual symptom combinations* approach. Originally designed to detect malingered emotional distress in individuals undergoing personal injury litigation, FBS has been shown to be useful in detecting somatic and cognitive overreporting (Ben-Porath, Graham, & Tellegen, 2009).

The MMPI-2 also includes scales intended to detect underreporting. Lie (L) was developed to assess defensiveness (underreporting). Developed using the *denial of minor flaws* method, it includes items that describe desirable but uncommon features. The *denial of psychopathology* method

---

[3] Despite the existence of several additional MMPI-2 validity scales in the literature, for brevity only those included in the standard test protocol are discussed.

was utilized to develop Correction (K), which was designed to distinguish between those who are genuinely asymptomatic and those who have symptoms but deny them. Butcher and Han (1995) used the *blended affirmation of virtuous behavior and denial of personal faults* strategy to develop the Superlative (S) scale, which includes items that differentiated between airline pilot applicants and individuals in the normative sample. Later research indicated that S is an effective indicator of underreporting (Ben-Porath, 2012).

Relatively few studies have examined the utility of MMPI-2 non–content-based invalid responding validity scales. Dragon, Ben-Porath, and Handel (2012) examined the impact of CNS elevations on the MMPI-2/MMPI-2-RF Restructured Clinical (RC) scales. They found the validity of substantive scales to be relatively robust to increasing missingness, but that the missingness tended to deflate scores to the point of changing the interpretation of test protocols. Several studies have demonstrated the sensitivity of VRIN to random responding (e.g., Berry et al., 1992; Gallen & Berry, 1996). Handel, Arnau, Archer, and Dandy (2006) demonstrated that TRIN is sensitive to simulated insertion of true or false item responses. A limitation of the extant literature has been the need to rely upon simulation designs to create missing, random, or fixed responses due to the difficulty in obtaining an external criterion of such response styles.

The MMPI-2 overreporting validity scales have been shown to distinguish between college students asked to respond honestly and those asked to overreport symptoms (Bagby, Buis, & Nicholson, 1995) as well as forensic pretrial defendants and psychiatric inpatients (Nicholson et al., 1997). They also appear to be effective in distinguishing between genuine psychopathology and simulated depression (Bagby, Marshall, & Bacchiochi, 2005) as well as simulated posttraumatic stress disorder (PTSD; e.g., Arbisi, Ben-Porath, & McNulty, 2006; Bury & Bagby, 2002; Elhai, Gold, Sellers, & Dorfman, 2001; Marshall & Bagby, 2006), and they show strong classification accuracy in distinguishing between Structured Interview of Reported Symptoms (SIRS)-classified overreporters and honest responders (Barber-Rioja, Zottoli, Kucharski, & Duncan, 2009). Based on meta-analytic findings, Rogers, Sewell, Martin, and Vitacco (2003) recommended the clinical use of F_p over F or F_B because the former had good classification rates and the latter appeared to be confounded by genuine psychopathology.

The MMPI-2 underreporting validity scales have been shown to distinguish between college student underreporting simulators and standard-instruction controls (e.g., Baer, Wetter, Nichols, Greene, & Berry, 1995; Bagby, Buis, & Nicholson, 1995). Baer and Miller's (2002) meta-analytic review of the MMPI-2 underreporting validity scale literature noted that several scales had moderately effective classification accuracy but were less effective in the presence of validity scale coaching. They recommended future research using known groups, differential prevalence, and

nonstudent simulation designs and examining the impact of coaching and the combined use of multiple MMPI-2 standard and supplementary under-reporting validity scales.

Minnesota Multiphasic Personality Inventory–2 Restructured Form

The 338-item MMPI-2 Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008) includes several revised MMPI-2 validity scales (CNS, VRIN-r, TRIN-r, F-r, F_p-r, FBS-r, L-r, K-r). Additionally, Infrequent Somatic Responses (F_S; Wygant, Ben-Porath, & Arbisi, 2004) was designed to identify individuals overreporting somatic complaints using the *rare symptoms* strategy and includes somatic items uncommonly endorsed by medical patients. The Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant, & Green, 2007) was empirically derived by selecting items that differentiated between disability claimants who failed and those who passed symptom validity tests.

Some research has examined the utility of this relatively new instru-ment in detecting non–content-based and content-based invalid respond-ing. Dragon et al. (2012) found that MMPI-2-RF CNS elevations changed the interpretability of substantive test results. Handel, Ben-Porath, Telle-gen, and Archer (2010) examined the effect of increasing levels of simu-lated random and fixed responding on VRIN-r and TRIN-r scores, find-ing that the scales were indeed sensitive to simulated non–content-based invalid responding. Ben-Porath (2012) reported that F-r is most sensitive to broad-range overreporting, while F_p-r is especially sensitive to overreported psychopathology. Sellbom and Bagby (2008b) found that L-r and K-r dis-tinguish between patients with schizophrenia who completed the test under standard or underreporting instructions. They found similar support for the scales in distinguishing between students who took the test under stan-dard instructions and (1) college student underreporting simulators as well as (2) child custody litigants. Although the extant literature is promising, more research is needed to examine the incremental utility and validity of the MMPI-2-RF validity scales across settings.

Personality Assessment Inventory

The Personality Assessment Inventory (PAI; Morey, 1991, 2007) is a 344-item personality and psychopathology self-report inventory with eight validity scales and indexes. Similar to the MMPI-2 VRIN scale, Incon-sistency (ICN) was designed to detect inconsistent responding using item pairs of similar or opposite content.

Infrequency (INF), designed to detect careless responding, consists of items that are not indicative of psychopathology but are so bizarre as to be rarely endorsed by individuals in normative and clinical samples. Nega-tive Impression (NIM), developed using the *rare symptoms* overreporting

detection strategy, consists of items with psychopathology or personal problem content that is unrealistically severe and not specific to any particular kind of psychopathology. The Malingering Index (MAL; Morey, 1993) was created using the *spurious patterns of psychopathology* strategy. Points are scored when an individual profile meets any of eight criteria that are more commonly endorsed by simulated overreporters than by honest responders (Morey, 1996). The Rogers Discriminant Function (RDF; Rogers, Sewell, Morey, & Ustad, 1996) was also developed using the *spurious patterns of psychopathology* strategy and includes indices that best distinguished between simulators instructed to feign specific disorders and those asked to respond honestly. Positive Impression Management (PIM) was developed utilizing the *denial of minor flaws* underreporting detection strategy, and is made of self-favorable items infrequently endorsed in normative and clinical samples. The Cashel Discriminant Function (CDF; Cashel, Rogers, Sewell, & Martin-Cannici, 1995) was developed utilizing an empirical approach that identified six scales (PIM and five substantive scales), which best distinguished between honest responding and simulated underreporting (Morey, 1996). The Defensiveness Index (DEF; Morey, 1993) is made of eight profile criteria, which occur more frequently with underreporting than with honest responding (Morey, 1996). DEF does not adhere to a particular underreporting detection strategy, but resembles the *blended affirmation of virtuous behavior and denial of personal faults* strategy (Sellbom & Bagby, 2008a). Recently, Mogge, Lepage, Bell, and Ragatz (2010) developed the Negative Distortion Scale (NDS) utilizing a *rare symptoms* approach, and Hopwood, Orlando, and Clark (2010) created the Malingered Pain-Related Disability-Discriminant Function (MPRDF) using a *spurious patterns of psychopathology* approach. Although initial results are promising, more research is needed to validate these validity indicators before they can be recommended for routine clinical use.

Sellbom and Bagby (2008a) reviewed the PAI literature on response distortion, noting that there was strong support for INF as an indicator of random responding, but that ICN had weaker classification accuracy results. They stated that INF and ICN were unable to identify partial random responding except at high levels. A new strategy developed by Morey and Hopwood (2004) shows promise in detecting back random responding, but more research is needed to replicate their findings.

Hawes and Boccaccini (2009) conducted a meta-analysis on the use of NIM, MAL, and RDF in the detection of overreported psychopathology, finding that all three differentiated between overreporters and standard-instruction responders with large effects. Further, research has demonstrated mean PAI validity scale score elevations in the overreporting of specific disorders, including psychosis, depression, generalized anxiety disorder, and posttraumatic stress disorder (e.g., Lange, Sullivan, & Scott, 2010; Rogers et al., 1996; Thomas, Hopwood, Orlando, Weathers, & McDevitt-Murphy, 2012). Of note, Sellbom and Bagby (2008a) concluded

that high false-positive rates indicate that the PAI may be effective at *screening out* but less effective at *screening in* malingering. Similarly, they concluded that the PAI demonstrates high false-positive rates in the detection of underreporting, but demonstrates promise as a screening tool.

The PAI validity scales were designed using a variety of detection strategies and are some of the most well-researched response distortion tools available. However, findings are mixed regarding the ability of the PAI to detect coached or sophisticated overreporting (Rogers et al., 1996; Thomas et al., 2012; Veltri & Williams, 2013) or underreporting (Baer & Wetter, 1997). Additional research on the impact of coaching and the incremental utility of each scale would be useful in further evaluating the utility of the PAI in multimethod assessment of response distortion (Sellbom & Bagby, 2008a).

Millon Clinical Multiaxial Inventory–III

The Millon Clinical Multiaxial Inventory–III (MCMI-III; Millon, 1994; Millon, Davis, & Millon, 1997), a self-report inventory designed to measure psychopathology and personality dysfunction in clinical settings, includes four validity indices. The Validity Index (Scale V) consists of three very *improbable symptoms* to assess for non–content-based invalid responding. The Disclosure Index (Scale X) was developed using the *indiscriminant symptom endorsement* strategy to detect whether the patient responded in an open and self-revealing or secretive manner. The Desirability Index (Scale Y) was developed using the *denial of minor flaws* strategy to measure "the patient's inclination to appear socially attractive, virtuous, or emotionally well composed" (Millon, 1994) and the Debasement Index (Scale Z) was designed using a simulated *rare symptoms* strategy (see Sellbom & Bagby, 2008a) to detect "an inclination to deprecate or devalue oneself by presenting more troublesome or emotional and personal difficulties than are likely to be uncovered upon objective review" (Millon, 1994).

Despite being developed to measure different aspects of invalid responding, X, Y, and Z appear to be intercorrelated and associated with MMPI-2 F and F_B, suggesting that all three are sensitive to quasi-rare overreported psychopathology (Craig, 1999; Morgan, Schoenberg, Dorr, & Burke, 2002). Scale Y appears to be modestly associated with MMPI-2 underreporting validity scales, indicating that it may be more sensitive to overreporting than to underreporting (Craig, 1999). These results were consistent with simulation studies that found X, Y, and Z to be associated with overreporting *and* underreporting (e.g., Bagby, Gillis, Toner, & Goldberg, 1991; Daubert & Metzler, 2000). Sellbom and Bagby (2008a) reviewed the MCMI-III literature, concluding that the extant literature on Scale Y indicated it was an inadequate measure of underreporting, with problematic classification accuracy results. They also noted that Scale X (intended to measure disclosure) may be a stronger indicator of overreporting than Scale

Z (intended to measure overreporting). They recommended against routine use of these scales because of problematic classification accuracy findings. In addition, they called for more cross-validation research as well as studies utilizing known groups and an examination of the effects of coaching.

Schedule for Nonadaptive and Adaptive Personality–2nd Edition

The Schedule for Nonadaptive and Adaptive Personality (SNAP; Clark, 1993) and its second edition (SNAP-2; Clark, Simms, Wu, & Casillas, in press) are self-report instruments designed to measure traits associated with personality disorders. Similar to the MMPI family, the SNAP instruments include VRIN and TRIN scales to measure variable and fixed non–content-based invalid responding, respectively. Desirable Response Inconsistency (DRIN) was designed to detect whether examinees inconsistently endorse socially desirable items while denying less socially desirable items of similar content. Conceptually similar to the MMPI-2 L Scale, Rare Virtues (RV) was designed using the *denial of personal faults* strategy and consists of rare and highly socially desirable items. Deviance (DEV) utilizes a *quasi-rare symptoms* approach to detect overreported problems. New to the SNAP-2, Back Deviance (BDEV) utilizes a *quasi-rare symptoms* approach to detect overreporting on the latter portion of the instrument.

VRIN, RV, DEV, BDEV, II, and to some degree, DRIN, are sensitive to extreme levels of non–content-based invalid responding. Most of the scales are sensitive to 100% acquiescent and 100% counteracquiescent responding as well as alternating response patterns (Clark et al., in press). Mean DEV and RV scores differ when comparing simulators and controls, with DEV evidencing strong sensitivity and specificity and RV evidencing strong specificity but modest sensitivity. Further, DEV appears to be associated with MMPI-2 F, whereas RV is associated with a variety of underreporting measures. At present, there is little support for the utility of DRIN, and users are urged to look for corroborating evidence of socially desirable responding (Clark et al., in press; Simms & Clark, 2001). Although initial studies show promise, more research is needed to validate the SNAP and SNAP-2 validity indices across settings.

Revised NEO Personality Inventory

The Revised NEO Personality Inventory (NEO PI-R; Costa & McCrae, 1992) is a 240-item measure of five personality domains. The authors included three one-item "Validity Checks" to screen out clearly invalid tests and guidelines for detecting random responding, but intentionally left out more comprehensive validity scales for the detection of underreporting or overreporting. Rather, they cautioned against administration of the self-report form "if the respondent is unlikely to understand the test

or is intensely motivated to present a false picture of himself or herself" (Costa & McCrae, 1992). In response to criticism regarding the lack of validity scales (e.g., Ben-Porath & Waller, 1992) and demonstrations that the NEO substantive scales are susceptible to response distortion (e.g., Paulhus, Bruce, & Trapnell, 1995), Schinka, Kinder, and Kremer (1997) developed three NEO PI-R validity scales: Inconsistency (INC; consisting of item pairs of similar content), Positive Presentation Management (PPM; consisting of quasi-rare positive qualities indicative of *social desirability*), and Negative Presentation Management (NPM; consisting of *quasi-rare symptoms* items) from the extant item pool. As these validity scales are neither part of the test report nor endorsed by either Pearson or the authors of the NEO PI-R, we do not review them here.

Standalone Measures for the Detection of Response Distortion

Structured Interview of Reported Symptoms

The Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992) is a 172-item structured interview designed to detect feigned psychopathology. The SIRS Primary Scales were designed using eight overreporting detection strategies and include: Rare Symptoms (RS), Symptom Combinations (SC), Improbable or Absurd Symptoms (IA), Blatant Symptoms (BL), Subtle Symptoms (SU), Severity of Symptoms (SEV), Selectivity of Symptoms (SEL), and Reported versus Observed Symptoms (RO). The test also includes a number of supplementary scales to provide additional information about the examinee's response style.

Research has demonstrated that SIRS factor scores are associated with MMPI-2 F and F_p scores (McCusker, Moran, Serfass, & Peterson, 2003). As presented in other sections, the SIRS has often been considered the overreporting gold standard, used to classify patients into criterion groups so that the utility of other tests (e.g., SIMS, M-FAST, MMPI-2, PAI) could be evaluated. To date, the most comprehensive analysis of the classification accuracy of the SIRS is a meta-analysis conducted by Green and Rosenfeld (2011), who examined SIRS studies and dissertations published from 1990 to 2009. The studies examined included original SIRS validation studies as well as replication analyses conducted after the test was published in 1992; the later studies demonstrated lower specificity but higher sensitivity values as compared to the original research. Composite effect sizes for the SIRS Total Score and averaged Primary Scales were notably large. The authors concluded that there is significant support for use of the SIRS, but that caution should be used in designating it as the gold standard of malingering detection, as other tools (e.g., MMPI-2) have demonstrated comparable utility and may be better suited for the detection of feigned cognitive deficits.

Structured Interview of Reported Symptoms, 2nd Edition

Four major changes were implemented in the development of the Structured Interview of Reported Symptoms, 2nd edition (SIRS-2; Rogers, Sewell, & Gillard, 2010): adding a classification scale (Rare Symptoms Total designed to differentiate between genuine but atypical and feigned presentations), two indices (Modified Total Index and Supplementary Scale Index), a supplementary scale for cognitive distortion (Improbable Failure), and a Decision Model to assist the clinician in making conclusions.

Since its release in 2010, the SIRS-2 and its manual have come under notable criticism. DeClue (2011) and Rubenzer (2010) noted such concerns as the use of only 36 suspected feigners in the development of the SIRS-2 decision rules, lack of information on the creation of criterion groups, questionable generalizability of the clinical normative sample (with half of patients diagnosed with dissociative identity disorder), and inflated sensitivity estimates due to the large number of indeterminate cases excluded from classification accuracy analyses. Further, DeClue (2011) reported that information comparing the classification accuracy of the SIRS and SIRS-2 on the same data is not available. Green, Rosenfeld, and Belfi (2013) compared SIRS and SIRS-2 scores in a criterion-group study of forensic inpatients and community overreporting simulators. The SIRS-2 tended to categorize more pretrial forensic patients as genuine or indeterminate responders as compared to SIRS-based classifications of the same scores. The SIRS-2 had excellent specificity but poor sensitivity, which was much lower than for the SIRS. They also found the SIRS Total Score was more useful than was the new SIRS-2 MT Index. More research is needed to inform clinicians about the utility this new instrument across a variety of settings.

Structured Inventory of Malingered Symptomatology

The Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005) is a 75-item self-administered inventory designed to measure overreported psychopathology and neuropsychological symptoms in clinical and forensic settings. The instrument has five scales designed to screen for various subdomains of overreported psychopathology, including Psychosis (P; bizarre psychotic symptoms not common in actual psychiatric patients), Neurologic Impairment (N; highly atypical or illogical neurological problems), Amnestic Disorders (Am; memory impairment not seen in individuals with actual brain injury), Low Intelligence (LI; simple, general fund of knowledge), and Affective Disorders (Af; atypical presentation of depression and anxiety). These scales were designed using a combination of detection strategies such as *improbable symptoms* and *close approximations to genuine symptoms* (Smith & Burger, 1997). Scores contribute to a Total Score that helps determine whether a more complete assessment of malingering is warranted (Smith, 2008).

The SIMS scales demonstrate expected associations with PAI and SIRS scales (Edens, Poythress, & Watkins-Clay, 2007). Further, SIMS is able to distinguish between simulated psychopathology overreporters and controls (e.g., Clegg, Fremouw, & Mogge, 2009; Edens, Otto, & Dwyer, 1999; Smith & Burger, 1997), as well as overreporting and honest groups as classified by SIRS scores (e.g., Alwes, 2006; Lewis, Simcox, & Berry, 2002) and by clinicians (Heinze & Purisch, 2001). Lewis et al. (2002) found that SIMS sensitivity and NPP were excellent, but specificity and PPP were low, supporting use of the SIMS as a screener that should be followed up with more extensive testing in the event of elevated scores. Presently, the body of literature on the SIMS is quite limited; more research would inform clinicians about its classification accuracy across settings.

Miller Forensic Assessment of Symptoms Test

The Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001) is a 25-item structured interview created to screen for overreported psychopathology in forensic settings. The instrument contains seven validity indices developed using a variety of overreporting strategies. The M-FAST scales include: Reported versus Observed (RO), Extreme Symptomatology (ES), Rare Combinations (RC), Unusual Hallucinations (UH), Unusual Symptom Course (USC), Negative Image (NI), and Suggestibility (S). The M-FAST scales demonstrate expected associations with MMPI-2, PAI, and SIRS scales (Gaines, 2009; Miller, 2001, 2004; Veazey, Hays, Wagner, & Miller, 2005). Several studies have indicated that the M-FAST is able to distinguish between simulated overreported psychopathology and honest responding (e.g., Guy et al., 2006) as well as between SIRS-classified overreporting and honest groups (e.g., Clark, 2006; Guy & Miller, 2004; Miller, 2004). Because the M-FAST is designed as a screener, it is recommended that a more extensive evaluation of overreporting be conducted if elevated scores occur (Miller, 2001). More research is needed on the utility and susceptibility to coaching of the M-FAST scales, especially those that consist of only one item.

Integrating Multiple Tools

Our aim in this chapter is to recommend sound practice in the multimethod assessment of response distortion. Earlier we reviewed extant research on the individual utility of a variety of standardized tools. Although not reviewed in detail, there is also a substantial literature on the incremental validity of multiple scales *within* many tests (e.g., incremental utility of MMPI-2 F_p over F).

Unfortunately, relatively few studies exist to inform clinicians of the incremental utility of multiple tests used together. Based on this limited

literature, it appears that the MMPI-2 and PAI validity indices explain more variance in combination than individually when distinguishing between overreporting simulators and psychiatric inpatients (Blanchard, McGrath, Pogge, & Khadivi, 2003) and between SIRS-identified honest and overreported protocols (Boccaccini, Murrie, & Duncan, 2006). The PAI Malingering Index (MAL) also appears to add incrementally to the M-FAST Total Score (Gaines, 2009). The SIRS adds a statistically significant but practically small increment over SIMS and PAI (Edens, Poythress, & Watkins-Clay, 2007) whereas the M-FAST incrementally improves upon the MMPI-2 in detecting overreporting (Clark, 2006).

Although less is known about the incremental utility of these instruments in detecting underreporting, moderate correlations between PAI and MMPI-2 underreporting validity indices (e.g., Weiss, Serafino, & Serafino, 2000) suggest that these inventories may be measuring unique domains of underreporting. More research is needed to understand whether the combined use of tools improves accuracy in the detection of underreporting.

Recommendations for Multimethod Assessment of Response Distortion

Response distortion is a multifaceted phenomenon that is affected by demand characteristics. Thus, examinees may alter their presentation across multiple evaluations. Further, the specific ways in which examinees will distort their responses is varied and can occur in three domains (psychopathology, cognitive, somatic). Because of the complexity of this phenomenon, we have several recommendations for the multimethod assessment of response distortion in personality and psychopathology evaluations.

• Gather background information that can inform you about potential motivating factors that can provide explanatory evidence regarding intentional or unintentional response distortion (e.g., Is there secondary gain potential? Has the examinee consistently presented in a self-deprecating manner with previous treatment providers?). Developing hypotheses about potentially motivating factors can help create pertinent interview questions and learn which domains of response distortion may be in need of greatest attention.

• Consider the evaluation to be an iterative process. Situational factors, referral questions, and initial information may inform the selection of interview questions, screening tools used, and collateral sources contacted. Initial test results may indicate the need for more comprehensive testing. Once sufficient evidence of response distortion is documented (utilizing validated measures with known error rates), further testing may be excessive and unnecessary. It is important to be prepared to alter planned test administration based on early results.

- Assess for reading level (when administering self-report instruments) and non–content-based invalid responding. Use interview responses and behavioral observations to bolster conclusions about the examinee's ability to attend to testing.

- At present, the current literature does little to inform clinicians about which *combination* of tools best detects various forms of response distortion. Thus, it may be best to select a variety of tools that have been individually validated in the examination setting of interest. Further, a combination should be selected so that at least one tool is sensitive to each type of response distortion (e.g., underreporting, general overreporting, and overreporting of specific symptoms). Further, it may be useful to select tools that were developed using a variety of detection strategies, as examinees may distort responses in a variety of ways. Because distortion often occurs across domains, at minimum screen for overreporting of cognitive and somatic complaints and refer for a neuropsychological evaluation if necessary.

- As recommended by Meehl (1955) and Ray (2009), use screening instruments in forensic settings in order to identify which examinees should be more comprehensively assessed. Screener cut scores are often set at low values that minimize false negatives at the cost of increased false positives. Thus, while the use of screeners (e.g., M-FAST, SIMS) may be cost-effective, it is important to further evaluate those who test positive with more comprehensive tests that utilize cut scores with lower levels of false positives.

- Select empirically supported tools. As shown in Table 12.2, validity indices measure different areas of response distortion. However, not all have been equally validated or examined across settings. Carefully select those that have been empirically validated in your setting and be aware of published error rates. Further, be cognizant that the base rate of invalid responding varies across settings; classification accuracy results from studies with unrealistic base rates may not generalize to your setting. If using substantive tools that do not have well-supported validity indices (e.g., Rorschach and BDI), also use well-validated validity indices in conjunction to assess for response bias.

- Be aware of the limitations to the DSM-5 Malingering V code. Although clinicians are tasked with using the problematic DSM-5 guidelines for malingering, it may be useful to conceptualize and assess for malingering using recommendations from Berry and Nelson (2010). Also be aware that it may be very difficult to gather information about motivating factors or intent. Thus, differentiating between malingering and other explanations (e.g., somatoform disorder, factitious disorder) without extensive historical information may present a significant challenge.

- Be mindful of the impact that culture may have on response styles. Research has demonstrated that individuals from different cultures tend to

vary in the extremity of their responses, level of acquiescence, and tendency to underreport problems (e.g., Aday, Cliu, & Anderson, 1980; Johnson, Kulesa, Cho, & Shavitt, 2005; Jürges, 2007; Mercado, 2000). It is important to gather information about the individual's cultural background to help determine whether their cultural norms may be impacting their presentation of symptoms.

Case Example

We selected a case demonstrating the importance of seven critical concepts in the assessment of response distortion.

Understand the Referral Question

Mr. Shaw is a 33-year-old male forensic hospital patient referred for an evaluation of competency to stand trial (CST) and malingering.

Thoroughly Review Available Background Information I: Previous Evaluations

Mr. Shaw was accused of committing three serious felonies, which, if he is convicted, could send him to prison for several years. He previously underwent seven competency evaluations and a hospital admission evaluation in the 21 months before his current evaluation. Some of the evaluators noted that Mr. Shaw reported a history of severe mental health symptoms, including believing he could speak with deceased people, appearing anxious, hypervigilant, agitated, paranoid, unable to sit still, and responding to internal stimuli. He also told an evaluator he was experiencing auditory hallucinations, including getting messages from the television to kill himself. Further, two evaluators noted that Mr. Shaw had difficulty recalling detailed information regarding his case and appeared confused. During one evaluation, he reported he could not remember the alleged offenses.

During other evaluations, it was noted that Mr. Shaw reported symptoms that appeared contrived (e.g., holding conversations with deceased people upon request by the examiner, delaying answers to give the impression of slowed cognitive processing). One examiner noted that he appeared conversational with peers but changed his mood to appear depressed and impaired when speaking with clinical staff. He was administered the M-FAST and received an overall score of 15, which was "highly suggestive of malingered psychopathology" (Miller, 2001). He did not demonstrate any genuine cognitive deficits. He provided significant historical information about his childhood, hospitalizations, and history of medication management and offered specific information about discussions with court personnel, including his status as an incompetent patient, options for plea deals, and the roles of court personnel. Mr. Shaw was able to state his

current charges and the severity of the charges, and to estimate the amount of incarceration time he could face if found guilty. He correctly identified his plea options, explained their meaning, and defined the court personnel, including their roles. He was able to think abstractly and had good insight, judgment, and impulse control.

In sum, the conclusions of previous evaluations were mixed. On several occasions, evaluators opined that Mr. Shaw's psychiatric and cognitive symptoms made him incompetent to stand trial, and he was subsequently transferred from the jail to a psychiatric hospital for competency restoration. Other evaluators opined that Mr. Shaw was competent to stand trial, and so he was returned to jail. Some evaluators suspected him of malingering. At the time of the current evaluation, he had recently been readmitted to the hospital, and his treatment team suspected he was malingering psychopathology and cognitive problems.

Thoroughly Review Available Background Information II: Other Records

Legal History

Records indicated that Mr. Shaw had an extensive history of juvenile and adult arrests and significant experience navigating the legal system and going to court.

Hospital Records Informing Mr. Shaw's Reported and Observed Cognitive Abilities

Hospital records indicated that Mr. Shaw was elected by his peers to be president of his unit, a coveted position that comes with special privileges and requires the ability to navigate the social milieu. The office is not typically obtained by patients who are especially suspicious of others, acutely psychotic, or severely depressed. Records also noted that Mr. Shaw attended court competency and discharge planning groups with little to no participation but took active part in leisure groups, demonstrating interest and being verbally expressive. It was noted that Mr. Shaw claimed to have memory problems, being unable to recall court procedure information. However, he was able to quickly and easily answer other questions about medication history, age of onset of illness, timeline of major events, and his daily schedule.

Hospital Records Informing Mr. Shaw's Reported and Observed Psychopathology

Mr. Shaw reported to staff that he heard voices, but he was never observed to be preoccupied with internal stimuli. He complained about previously being assigned a Malingering V code, but he also admitted that he had fabricated psychotic symptoms during his first admission to the hospital.

Conduct the Clinical Interview

While the purpose of this evaluation was to understand Mr. Shaw's current competency to stand trial, it was essential to ask about his reported history of cognitive and psychopathology symptoms to understand whether he had genuine problems that interfered with his ability to attend to court proceedings and assist his attorney.

Self-Reported Cognitive Deficits

Mr. Shaw reported a history of concussions as a teenager. When asked whether he experienced memory deficits related to his case, he reported he did not know what a jury was or whether it included more than one person. He said he did not know what made him the defendant but that he had "heard of" a judge. He said his public defender is "someone who tries to convict you" and said he is incompetent "because I forget. I don't know all of the material yet."

Self-Reported Psychiatric Symptoms

Mr. Shaw reported that he began hearing voices around age 8 and hears them daily. He said he was previously diagnosed with bipolar disorder and has a history of suicide attempts. In describing his recent symptoms, Mr. Shaw reported, "I thought the government was after my family. Most of the time, I believe it. I can hear agents' names, and I trip out."

Self-Reported Information Regarding Potential Motivation to Distort

Mr. Shaw said he prefers the hospital to the jail because "It's a less stressful environment . . . than being locked in a cell 24 hours a day." He reported he receives better treatment, proper medical care, and therapy at the hospital. He also talked at length about a hospital peer whom he identified as his girlfriend.

Pay Attention to Behavioral Observations

It can help to examine whether an examinee's behaviors are inconsistent with his or her reported symptoms. Such information can be invaluable in determining whether they are reporting false symptoms. For Mr. Shaw, the following behavioral observations were noted:

Cognitive Abilities

A day after the evaluator held a brief conversation with Mr. Shaw to meet him and schedule a time for testing, Mr. Shaw saw the evaluator on hospital

grounds. Without prompting, he said, "Tomorrow at 9:30, right?" indicating recognition of an unfamiliar staff member and recall of the time of the scheduled appointment. Despite answering many questions with "I don't know," Mr. Shaw appeared to be very cognizant of the tests being administered. During administration of the M-FAST, he correctly stated, "I already did this one when I got here," indicating he recalled the front page of the test from an evaluation conducted several months earlier. Mr. Shaw was able to attend to items, follow test instructions independently, and maintain focus on test items for up to 90 minutes at a time. He demonstrated good memory, attention, and concentration. In contrast to interview responses that, on the surface, indicated he had little knowledge of the legal system, Mr. Shaw provided significant detail about previous arrests, the crimes committed that led to them, and the negotiation process he experienced for previous plea deals.

Psychopathology

Mr. Shaw reported hearing voices talking and repeating the evaluator's questions during the current interview. However, at no point did he appear to be distracted by internal stimuli during several hours of the evaluation. Despite reporting highly persecutory delusional content about legal players wishing to unjustly give him the death penalty, steal his money, and kill his family, he slouched in his chair and answered questions with his arms crossed and in a relaxed position. He presented as euthymic throughout the interview with no signs of anxious, depressive, or manic symptomatology.

Decide upon Instruments to Administer

Previous evaluators came to a variety of conclusions about the veracity of Mr. Shaw's symptomatology. The only standardized testing available for review was an M-FAST administered during a hospital intake. While his very high score of 15 provided evidence of overreported psychopathology, the M-FAST is a state-dependent screening instrument and further evaluation was necessary to more comprehensively assess for distortion. Further, despite reported memory difficulties, no previous testing assessed for cognitive response distortion.

To accurately assess barriers to Mr. Shaw's current competence to stand trial, it was necessary to assess whether he was *currently* exaggerating or fabricating cognitive or mental health problems. There was no indication of need to comprehensively assess for overreported somatic symptoms or underreporting. As listed in Table 12.3, several detection methods were chosen within the selected domains because feigners may distort responses using several approaches and different tests are sensitive to different distortion approaches. Of note, test and interview question selection is an

TABLE 12.3. Instruments Administered to Mr. Shaw and His Test Results

Name of instrument	Description of instrument	Mr. Shaw's scores	Basic interpretation of scores
		Cognitive domain	
Rey Fifteen Item Memorization Test (FIT)	Screener measure of suspect effort on cognitive tests	Combined Recognition and Recall Score: 23	There was no evidence of suspect effort.
Test of Memory Malingering (TOMM)	Forced-choice screener of suspect effort on memory and other cognitive tests	Trial 1: 41/50 Trial 2: 40/50 Retention: 27/50	Mr. Shaw's test scores indicate he did not put forth maximum effort and may have attempted to appear as though he had memory deficits that were not true for him.
Validity Indicator Profile (VIP)	Forced-choice measure of effort and intent on cognitive tests	Nonverbal: 88; Compliant Verbal: 67; Compliant	Mr. Shaw had a compliant response style. There was no compelling evidence of suppression, random responding, or pronounced inconsistency.
Inventory of Legal Knowledge (ILK)	Forced-choice measure of response style in defendants undergoing adjudicative competence evaluations	28/61 (45.9%)	Mr. Shaw's score of 28/61 was similar to scores obtained by examinees who respond randomly. It was lower than scores obtained by most examinees with bona fide mental disorders. His score suggests he had little investment in demonstrating his true knowledge or abilities and raises significant concerns about a feigned or irrelevant response style.
		Psychopathology domain	
Miller Forensic Assessment of Symptoms Test (M-FAST)	Screener structured interview designed to quickly screen for psychopathology overreporting	Previous: 15 Current: 8 RO: 1	Mr. Shaw endorsed items on the M-FAST that indicate the possibility of malingered mental illness. An M-FAST score of 8 is highly suggestive of malingered mental

368

ES: 2
RC: 3
UH: 2

illness. A M-FAST score of 8 is highly suggestive of malingered psychopathology in clinical samples. The Reported versus Observed (RO) Scale includes items indicating self-reported problems are inconsistent with observed behaviors. The Extreme Symptomatology (ES) Scale includes symptoms that are very extreme and uncommon. The Rare Combinations (RC) Scale includes psychological symptoms that are rarely seen in combination. The Unusual Hallucination (UH) Scale includes symptoms that are extremely rare in genuine psychiatric populations. Based on Mr. Shaw's M-FAST scores, additional tests are recommended in order to evaluate the possibility that he is malingering.

Structured Interview of Reported Symptoms (SIRS)

Comprehensive structured interview designed to assess psychopathology overreporting response styles (and, less comprehensively, defensiveness)

1 Definite (Rare Symptoms) and 3 Probable (Blatant Symptoms, Selectivity of Symptoms, and Severity of Symptoms)

Mr. Shaw had a markedly elevated score on Rare Symptoms, which measures the tendency to endorse very severe symptoms, often with psychotic content. This elevation is characteristic of individuals who are feigning a mental disorder and is rarely seen in clients responding truthfully. He also had moderately elevated scores on Blatant Symptoms (endorsing many symptoms associated with major mental illness), Selectivity of Symptoms (endorsing a wide range of psychiatric symptoms), and Severity of Symptoms (endorsing many symptoms as extreme or unbearable). This combination of elevated scores is characteristic of individuals who are feigning a mental disorder and is rarely seen in clients responding truthfully.

iterative process. Mr. Shaw's behaviors and test results earlier in the assessment process influenced the selection of later questions and tests. Table 12.3 provides Mr. Shaw's scores along with basic interpretive results for each test. We will not devote a great deal of attention to Mr. Shaw's results in the cognitive symptoms domain (which suggested he did not attempt to perform as well as he could). Rather, we will turn our attention to Mr. Shaw's results from the psychopathology domain.

Mr. Shaw was first administered the Miller Forensic Assessment of Symptoms Test (M-FAST). While his current score of 8 was much lower than his previous score of 15, it still suggested overreported psychopathology symptoms. Because the M-FAST was designed as a brief screening tool, Mr. Shaw was also administered the Structured Inventory of Reported Symptoms (SIRS). He scored in the "Definite" range of overreporting on one of the eight primary scales and in the "Probable" range of overreporting on three others. This provides more information about his approach, utilizing a more comprehensive measure of psychopathology response distortion. Although not administered during this evaluation, we might have additionally chosen to administer a self-report multiaxial personality instrument (e.g., MMPI-2-RF, PAI) which would utilize a different approach. We refrained from doing so because we had significant evidence of overreported psychopathology and did not wish to subject the patient to unnecessary testing.

Interpret Multimethod Test Results with Behavioral Observations and Records

Based on the available evidence, we concluded that Mr. Shaw was intentionally fabricating or exaggerating cognitive difficulties and psychopathology in an attempt to remain at the hospital. This conclusion can be broken into three main areas of evidence. Of note, not all three will be clearly determined in every evaluation:

Behavior

Mr. Shaw's performance on measures of distortion of cognitive problems, as well as discrepancies between his behavior during testing and when not formally evaluated, indicate that he was overreporting psychopathology symptoms and putting forth less than maximal effort to perform well on cognitive measures. He previously demonstrated good knowledge of the court process, which was much better than his reported level of knowledge during this evaluation. It is doubtful that he forgot such information considering that his memory was observed to be intact and there were no documented physiological or psychiatric problems that would indicate he had forgotten the legal information he previously knew.

Intent to Distort

According to records, Mr. Shaw admitted to intentionally overreporting psychopathology in the past. His test results indicated that he was likely overreporting psychiatric problems and not performing as well as he could on cognitive measures. However, *Mr. Shaw's response distortion test results did not directly speak to his intent.* We had to infer his intent based on discrepancies between his behavior during and outside of formal evaluation as well as discrepancies in demonstrated cognitive abilities.

Motivation for Behavior

According to records and his self-report during this evaluation, Mr. Shaw reported a desire to remain at the hospital as opposed to jail. This (admittedly rare) admission allowed for direct conclusions about his motivation to distort. Less direct evidence (e.g., knowledge that he enjoyed spending time with his girlfriend) or hypotheses (e.g., his desire to delay the trial to avoid being found guilty) may have been *inferred,* but without more direct evidence, such inferences would not confirm his motivation.

Although it could not be concluded certainly as to whether Mr. Shaw had any bona fide psychiatric symptoms, his presentation suggested that he was grossly overreporting symptoms of mental illness and memory deficits due to external incentives. As such, we opined that he was competent to stand trial. Further, we recommended that his treatment team consider adding a Malingering V code to his diagnosis of record.

Conclusions

As highlighted throughout this chapter and case example, the assessment of response distortion is a multifaceted endeavor that involves consideration of information from a variety of sources. This task is made more complex by imperfect guidelines (e.g., DSM-5), which have been described as incomplete and unattainable (e.g., Berry & Nelson, 2010). Fortunately, several well-validated methods exist to guide clinicians willing to take on this challenge. Careful selection of standardized tools in conjunction with information from interview, behavioral observations, and collateral sources can provide rich and useful information about the presence, severity, and type of response distortion in order to inform clinical and forensic practice.

REFERENCES

Aday, L. A., Cliu, G. Y., & Anderson, R. (1980). Methodological issues in health care surveys of the Spanish-descent population. *American Journal of Public Health, 70,* 367–374.

responding on the Millon Clinical Multiaxial Inventory–III. *Psychological Assessment, 12,* 418–424.

DeClue, G. (2011). Harry Potter and the Structured Interview of Reported Symptoms? *Open Access Journal of Forensic Psychology, 3,* 1–18.

Dragon, W. R., Ben-Porath, Y. S., & Handel, R. W. (2012). Examining the impact of unscorable item responses on the validity and interpretability of MMPI-2/ MMPI-2-RF Restructured Clinical (RC) scale scores. *Assessment, 19,* 101–113.

Edens, J. F., Otto, R. K., & Dwyer, T. (1999). Utility of the Structured Inventory of Malingered Symptomatology in identifying persons motivated to malinger psychopathology. *Journal of the American Academy of Psychiatry and the Law, 27,* 387–396.

Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment, 88,* 33–42.

Edwards, A. L. (1970). *The measurement of personality traits by scales and inventories.* New York: Holt, Rinehart, & Winston.

Elhai, J. D., Gold, S. N., Sellers, A. H., & Dorfman, W. I. (2001). The detection of malingered posttraumatic stress disorder with MMPI-2 fake bad indices. *Assessment, 8,* 221–236.

Gaines, M. V. (2009). *An examination of the combined use of the PAI and the M-FAST in detecting malingering among inmates.* Unpublished doctoral dissertation. Texas Tech University, Lubbock, TX.

Gallen, R. T., & Berry, D. T. R. (1996). Detection of random responding in MMPI-2 protocols. *Assessment, 3,* 171–178.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14,* 196–208.

Graham, J. R. (2006). *MMPI-2: Assessing Personality and Psychopathology* (4th ed.). New York: Oxford University Press.

Green, D., & Rosenfeld, B. (2011). Evaluating the gold standard: A review and meta-analysis of the Structured Interview of Reported Symptoms. *Psychological Assessment, 23,* 95–107.

Green, D., Rosenfeld, B., & Belfi, B. (2013). New and improved?: A comparison of the original and revised versions of the Structured Interview of Reported Symptoms. *Assessment, 20,* 210–218.

Guy, L. S., Kwartner, P. P., & Miller, H. A. (2006). Investigating the M-FAST: Psychometric properties and utility to detect diagnostic specific malingering. *Behavioral Sciences and the Law, 24,* 687–702.

Guy, L. S., & Miller, H. A. (2004). Screening for malingered psychopathology in a correctional setting: Utility of the Miller-Forensic Assessment of Symptoms Test (M-FAST). *Criminal Justice and Behavior, 31,* 695–716.

Hall, H. V., & Poirier, J. G. (2001). *Detecting malingering and deception* (2nd ed.). Boca Raton, FL: Taylor & Francis.

Handel, R. W., Arnau, R. C., Archer, R. P., & Dandy, K. L. (2006). An evaluation of the MMPI-2 and MMPI—A True Response Inconsistency (TRIN) scales. *Assessment, 13,* 98–106.

Handel, R. W., Ben-Porath, Y. S., Tellegen, A., & Archer, R. P. (2010). Psychometric functioning of the MMPI-2-RF VRIN-r and TRIN-r scales with varying degrees of randomness, acquiescence, and counter-acquiescence. *Psychological Assessment, 22,* 87–95.

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology

on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment, 21,* 112–124.

Heinze, M. C. (1999). "Yet there's method in his madness . . .": Dimensions of deception and dangerousness. *Aggression and Violent Behavior, 4,* 387–412.

Heinze, M. C., & Purisch, A. D. (2001). Beneath the mask: Use of psychological tests to detect and subtype malingering in criminal defendants. *Journal of Forensic Psychology, 1,* 23–52.

Hoelzle, J. B., Nelson, N. W., & Arbisi, P. A. (2012). MMPI-2 and MMPI-2-Restructured Form validity scales: Complimentary approaches to evaluate response validity. *Psychological Injury and Law, 5,* 174–191.

Hopwood, C. J., Orlando, M., & Clark, T. C. (2010). The detection of malingered pain-related disability with the Personality Assessment Inventory. *Rehabilitation Psychology, 55,* 307–310.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology, 36,* 264–277.

Jürges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics, 16,* 163–178.

Kucharski, L. T., Duncan. S., Egan, S. S., & Falkenbach, D. M. (2006). Psychopathy and malingering of psychiatric disorder in criminal defendants. *Behavioral Sciences and the Law, 24,* 633–644.

Lange, R. T., Sullivan, K. A., & Scott, C. (2010). Comparison of MMPI-2 and PAI validity indicators to detect feigned depression and PTSD symptom reporting. *Psychiatry Research, 176,* 229–235.

Larrabee, G. L. (Ed.). (2007). *Assessment of malingered neuropsychological deficits.* New York: Oxford University Press.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal-injury claimants. *Psychological Reports, 68,* 203–201.

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,* 170–176.

Marshall, M., & Bagby, R. M. (2006). The incremental validity and clinical utility of the MMPI-2 Infrequency Posttraumatic Stress Disorder scale. *Assessment, 13,* 417–429.

McCusker, P. J., Moran, M. J., Serfass, L., & Peterson, K. H. (2003). Comparability of the MMPI-2 F(p) and F scales and the SIRS in clinical use with suspected malingerers. *International Journal of Offender Therapy and Comparative Criminology, 47,* 585–596.

Meehl, P. E. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194–216.

Meehl, P. E. (1996). Preface. In *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Northvale, NJ: Jason Aronson. (Original work published 1954)

Mercado, M. M. (2000). The invisible family: Counseling Asian American substance abusers and their families. *Family Journal: Counseling and Therapy for Couples and Families, 8,* 267–272.

Mihura, J. L. (2012). The necessity of multiple test methods in conducting assessments: The role of the Rorschach and self-report. *Psychological Injury and Law, 5,* 97–106.

Miller, H. A. (2001). *Miller-Forensic Assessment of Symptoms Test (M-FAST): Professional manual.* Odessa, FL: Psychological Assessment Resources.

Miller, H. A. (2004) Examining the use of the M-FAST with criminal defendants incompetent to stand trial. *International Journal of Offender and Comparative Criminology, 48,* 268–280.

Millon, T. (1994). *Millon Clinical Multiaxial Inventory–III manual.* Minneapolis, MN: National Computer Systems.

Millon, T., Davis, R., & Millon, C. (1997). *Millon Clinical Multiaxial Inventory–III manual* (2nd ed.). Minneapolis, MN: National Computer Systems.

Mogge, N. L., Lepage, J. S., Bell, T., & Ragatz, L. (2010). The negative distortion scale: A new PAI validity scale. *Journal of Forensic Psychiatry and Psychology, 21,* 77–90.

Morey, L. C. (1991). *Personality Assessment Inventory professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (1993, August). *Defensiveness and malingering indices for the PAI.* Paper presented at the annual convention of the American Psychological Association, Toronto, Ontario, Canada.

Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory.* Odessa, FL: Personality Assessment Resources.

Morey, L. C. (2007). *Personality Assessment Inventory professional manual,* 2nd ed. Odessa, FL: Psychological Assessment Resources.

Morey, L. C., & Hopwood, C. J. (2004). Efficiency of a strategy for detecting back random responding on the Personality Assessment Inventory. *Psychological Assessment, 16,* 197–200.

Morgan, C. D., Schoenberg, M. R., Dorr, D., & Burke, M. J. (2002). Overreport on the MCMI-III: Concurrent validation with the MMPI-2 using a psychiatric inpatient sample. *Journal of Personality Assessment, 78,* 288–300.

Morgan, J. E., & Sweet, J. J. (Eds.). (2009). *Neuropsychology of malingering casebook.* New York: Psychology Press.

Nicholson, R. A., Mouton, G. J., Bagby, R. M., Buis, T., Peterson, S. A., & Buigas, R. A. (1997). Utility of MMPI-2 indicators of response distortion: Receiver operating characteristic analysis. *Psychological Assessment, 9,* 471–479.

Paulhus, D. L. (1988). *Assessing self-deception and impression management in self-reports: The Balanced Inventory of Desirable Responding.* Unpublished manual, University of British Columbia, Vancouver, Canada.

Paulhus, D. L., Bruce, M. N., & Trapnell, P. D. (1995). Effects of self-presentation strategies on personality profiles and their structure. *Personality and Social Psychology Bulletin, 21,* 100–108.

Ray, C. L. (2009). The importance of using malingering screeners in forensic practice. *Journal of Forensic Psychological Practice, 9,* 138–146.

Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law, 2,* 93–112.

Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice, 21,* 182–188.

Rogers, R. (2008a). An introduction to response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 3–13). New York: Guilford Press.

Rogers, R. (2008b). Detection strategies for malingering and defensiveness. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 14–38). New York: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms professional manual.* Odessa, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In I. B. Weiner (Series Ed.) & A. M. Goldstein (Vol. Ed.), *Handbook of psychology: Vol. 11. Forensic psychology* (pp. 109–132). Hoboken, NJ: Wiley.

Rogers, R., & Cavanaugh, J. L. (1983). "Nothing but the truth" . . . A reexamination of malingering. *Journal of Law and Psychiatry, 11,* 443–460.

Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured Interview of Reported Symptoms* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

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., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment, 67,* 629–640.

Rubenzer, S. (2010). Review of the Structured Inventory of Reported Symptoms–2 (SIRS-2). *Open Access Journal of Forensic Psychology, 2,* 273–286.

Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. *Journal of Personality Assessment, 68,* 127–138.

Sellbom, M., & Bagby, R. M. (2008a). Response styles on multiscale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 182–206). New York: Guilford Press.

Sellbom, M., & Bagby, R. M. (2008b). Validity of the MMPI-2-RF (Restructured Form) L-r and K-r scales in detecting underreporting in clinical and nonclinical samples. *Psychological Assessment, 20,* 370–376.

Simms, L. J., & Clark, L. A. (2001). Detection of deception on the Schedule for Nonadaptive and Adaptive Personality: Validation of the validity scales. *Assessment, 8,* 251–266.

Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 323–342). New York: Guilford Press.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25,* 183–189.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review, 9,* 219–242.

Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, N. (1999). Taxometric analysis of impression management and self-deception on the MMPI-2 in child custody litigants. *Journal of Personality Assessment, 73,* 1–18.

Tellegen, A. (1985). Structure of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.

Tett, R. P., Anderson, M. G., Ho, C.-L., Yang, T. S., Huang, L., & Hanvongse, A. (2006). Seven nested questions about faking in personality tests. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior.* Greenwich, CT: Information Age.

Thomas, K. M., Hopwood, C. J., Orlando, M. J., Weathers, F. W., & McDevitt-Murphy, M. E. (2012). Detecting feigned PTSD using the Personality Assessment Inventory. *Psychological Injury and Law, 5,* 192–201.

Thomas-Peter, B. A., Jones, J., Campbell, S., & Oliver, C. (2000). Debasement and

faking bad on the Millon Clinical Multiaxial Inventory III: An examination of characteristics, circumstances and motives of forensic patients. *Legal and Criminological Psychology, 5,* 71–81.

Veazey, C. H., Hays, J. R., Wagner, A. L., & Miller, H. A. (2005). Validity of the Miller Forensic Assessment of Symptoms Test in psychiatric inpatients. *Psychological Reports, 96,* 771–774.

Veltri, C. O. C., & Williams, J. E. (2013). Does the disorder matter?: Investigating a moderating effect on coached malingering using the MMPI-2 and PAI. *Assessment, 20,* 199–209.

Vitacco, M. J. (2008). Syndromes associated with deception. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 39–50). New York: Guilford Press.

Weiss, W. U., Serafino, G., & Serafino, A. (2000). A study of the interrelationships of several validity scales used in police selection. *Journal of Police and Criminal Psychology, 15,* 41–44.

Widows, M., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology (SIMS): Professional manual.* Odessa, FL: Psychological Assessment Resources.

Wiggins, C. W., Wygant, D. B., Hoelzle, J. B., & Gervais, R. O. (2012). The more you say the less it means: Overreporting and attenuated criterion validity in a forensic disability sample. *Psychological Injury and Law, 5,* 162–173.

Wygant, D. B., Ben-Porath, Y. S., & Arbisi, P. A. (2004). *Development and initial validation of a scale to detect infrequent somatic complaints.* Poster presented at the 39th annual symposium on recent developments of the MMPI-2/MMPI-A, Minneapolis, MN.