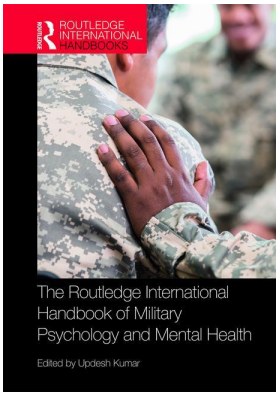

The Routledge International Handbook of Military Psychology and Mental Health

Updesh Kumar

Validity assessment in military psychology

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9780429281266-15
Noah K. Kaufman, Shane S. Bush
Published online on: 19 Dec 2019

15

VALIDITY ASSESSMENT IN MILITARY PSYCHOLOGY

Noah K. Kaufman and Shane S. Bush

Military personnel are at risk for cognitive, emotional, and behavioral problems. Whether in training, routine daily activities, or combat environments, the potential exists for physical or psychological trauma. Psychological and neuropsychological evaluations, therefore, serve an essential role in understanding the possible impact of such events on the functioning and wellbeing of military personnel. For the collected data to reflect the constructs of interest (e.g., anxiety, memory), test-takers must respond accurately to questions about their backgrounds, put forth good effort on cognitive ability tests, and respond in a consistent and honest manner to inventories of psychological functioning. Failure to respond in a valid manner limits the usefulness of the results for the purposes for which they were intended, except when the purpose of the evaluation is to assess for the presence of deceptive behavior (e.g., malingering).

Military personnel may generate invalid data for multiple reasons; for example, they may minimize problems in order to remain with or return to their units, or they may exaggerate or fabricate problems to avoid unpleasant or dangerous duties or to obtain benefits. Accordingly, a multi-method approach to assessing the validity of symptoms and performance is an essential component of psychological and neuropsychological evaluations of military personnel, as well as veterans.

The purpose of the present chapter is to provide an evidence-based review of validity assessment, including discussion of common biasing factors operating *during* and *after* data collection; symptom, performance, and response validity; and misconceptions about validity assessment. Because the availability and use of standardized assessment measures varies within and among countries and cultures, an effort has been made to emphasize principles and procedures that are broadly relevant and therefore more readily useful to practitioners across a range of specific contexts.

Psychological assessment in the military

In the United States, the field of psychological assessment got much of its impetus from World War I (Yoakum & Yerkes, 1920) because it was necessary "to point out the feeble-minded and those incapable of military service because of mental deficiency and … to find those of unusual or special ability" (p. 10). Test development and validation during this early period included

uncommonly large sample sizes (N = 1,726,966), and test-taking instructions were reflective of the times (Yoakum & Yerkes, 1920):

> Attention! The purpose of this examination is to see how well you can remember, think, and carry out what you are told to do. We are not looking for crazy people. The aim is to help find out what you are best fitted to do in the Army. The grade you make in this examination will be put on your qualification card and will also go to your company commander. Some of the things you are told to do will be very easy. Some you may find hard. You are not expected to make a perfect grade, but do the very best you can. (p. 53)

At present, aptitude testing remains an integral part of psychological assessment in the military with new recruits. However, it is also currently important for mental health professionals to accurately characterize mental health problems among military personnel and veterans. Because both types of assessment, cognitive and general mental health, routinely take place in the military; validity assessment is also routinely needed.

Deception and validity assessment in the military

Deception can be an understandable, and effective, strategy to improve one's circumstances. Animals use deception for this reason (Stone & Boone, 2007), as do humans, including very young children (Peterson & Peterson, 2015). The implication here is that one need not be particularly experienced or sophisticated to intentionally deceive.

In his discussion of this topic, Frederick (2012) emphasized that intentional deception to achieve a desired outcome is not inherently bad and that it can sometimes be understandably adaptive: "It is therefore usually not useful to consider people 'bad' who malinger. In the right context, malingering can be considered heroic behavior—a way to deceive enemies. King David, for example, once escaped from his enemies by feigning mental illness" (p. 229). That said, there are times where the intentional use of deception is not heroic, in which case strategies to correctly identify deceivers are needed. For example, deception by criminal defendants to improve their legal circumstances can be a cowardly act to evade punitive servitude.

Another form of servitude, although not of a penal nature, is military service. In the military, deception through exacerbation or fabrication of problems is used to avoid service and to obtain medical discharges and/or disability-based benefits. However, the percentage, or base rate, of deceivers in the military has not been well established. To illustrate, in a recent large-scale study with a sample size of 28,065,568 military healthcare visits, clinicians only diagnosed 1,074 individuals as malingering or having factitious disorder (Lande & Williams, 2013). Other estimates put the base rate at about 0.5% (Armed Forces Health Surveillance Center [AFHSC], 2013). If one accepts these numbers, deception to improve one's circumstances in the military is practically non-existent.

In contrast, Denning and Shura (2017) found that 33% to 52% of disability-seeking military veterans were malingering mild traumatic brain injury (mTBI), which does not take into consideration other forms of malingering. Armistead-Jehle (2010) reported very high rates (58%) of inadequate effort output among veteran examinees referred for mTBI-related evaluations, a theme supported by data from Wall, Graver, and Shurak (2013), who found that over 50% failed validity testing in their sample of 1,178 active-duty soldiers, military retirees, and military beneficiaries. Meanwhile, Whitney, Shepard, Williams, Davis, and Adams (2009) reported lower rates (17%) of veterans from either Operation Iraqi Freedom or Operation Enduring Freedom scoring below cut-offs on validity testing. Similarly, over half of American Vietnam veterans in

a study by Freeman, Powell, and Kimbrell (2008) exaggerated symptoms of posttraumatic stress disorder (PTSD).

Because studies examining exaggeration of problems do not cover military personnel who deny or minimize symptomatology for the purpose of returning to, or remaining with, their unit or for other psychological reasons, the full extent of deception in the military remains unknown. Moreover, because currently reported estimates of intentionally deceptive behavior in the military vary so radically, it is premature to conclude that a solid understanding as to the base rate for this behavior exists.

It is not premature, however, to appreciate that estimates of malingering in the military are influenced by two factors: (1) the negative consequences of diagnosing malingering *in the military* and (2) the methods used to identify malingerers. Regarding the first factor, those identified as malingering may be tried in military court (i.e., court-martialed) under Article 115 of the Uniform Code of Military Justice (UCMJ). Although there are exceptions in non-military settings (e.g., *U.S. v. Greer, 1998*), non-military malingering generally does not trigger, or enhance, legal consequences. Regarding the second factor, many healthcare clinicians working in a military context do not use scientific methods to detect malingering, which is problematic given what has been known about the superiority of actuarial methods, as opposed to unstructured clinical judgment, since the 1950s (Meehl, 1954) and that methods for measuring deception have existed for decades (Pankratz, 1979; Rey, 1941). While the field of neuropsychology fully embraces the importance of validity assessment (Bush et al., 2005), the same does not hold for other healthcare fields.

In short, some mental health clinicians working with military populations are timid about labeling soldiers and veterans as malingerers because of harsher consequences, which affects base rate estimates. Similarly, mental health clinicians often do not go beyond unstructured clinical judgment to detect deception, also skewing base rate estimates.

Biasing factors influencing psychological and neuropsychological test scores

Psychological and neuropsychological assessment in any context involves the collection of data. These data are vulnerable to many biasing influences, as conveyed in Figure 15.1. It is, therefore, always necessary for psychologists to consider these influences and, sometimes, account for their impact on the data.

The influence of demographic variables such as sex, education level, racial group, and age can be statistically controlled for with many neuropsychological tests, the importance of which has been articulated in the neuropsychological literature. In contrast, other tests (e.g., most measures of IQ, academic achievement, personality) tend to use stratified general U.S. population norms (Strauss, Sherman, & Spreen, 2006), meaning the examinee's score is computed in comparison to others their age and what is normal for all others in the entire United States, across levels of sex, education level, racial group, and socioeconomic status.

Calibrating scores based on multiple demographic factors is not always necessary because some referral questions call for a comparison of the examinee to what is normal in the general population. Although it is typically necessary to control for age, it is not always necessary to control for other demographic variables, especially if the demographic variable does not confound the neuropsychological data point of interest (e.g., memory, attention, executive functioning). If the demographic variable *does* confound the neuropsychological data point, it is then necessary to consider the nature of the relationship. To illustrate, age, sex, and racial group status can have a *unidirectional* relationship with neuropsychological data points, whereas education level has a *bidirectional* relationship with neuropsychological data points (Boone, 2013a), meaning that

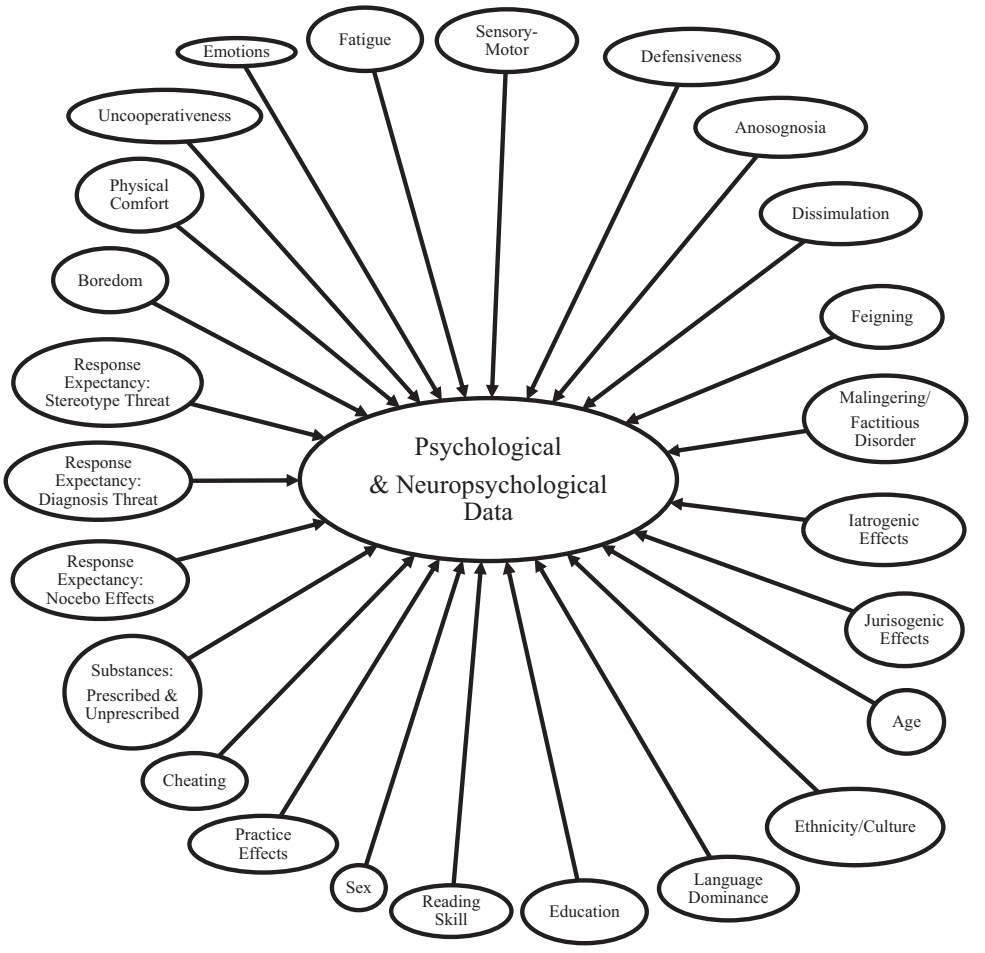

Figure 15.1 Biasing factors influencing psychological and neuropsychological data.

cognitive ability influences how much education a person gets, and how much education a person gets influences their cognitive abilities.

Regarding the manner in which examinees respond to questions from psychological and neuropsychological tests (i.e., their *response style*), different terminology is encountered. Rogers and Bender (2013) provided helpful definitions, the following of which are included in Figure 15.1:

> *malingering* ("deliberate fabrication or gross exaggeration of psychological or physical symptoms for the fulfillment of an external goal"); *feigning* ("deliberate fabrication or gross exaggeration of psychological or physical symptoms without any assumptions about goals [for this behavior]"; *dissimulation* ("general term to describe an inaccurate portrayal of symptoms … used when more precise terms [do not clearly apply]"); and *defensiveness* ("the polar opposite of malingering; it is the deliberate denial or gross minimization of symptoms in the service of an external goals"). (p. 518)

Other response styles in Figure 15.1 are *uncooperativeness* (Vore, 2007), which is when an examinee is unwilling to go through with all or part of the evaluation, and *factitious disorder*, which is

fabricated or exaggerated health problems for the purpose of receiving attention from healthcare providers. Other terms not included in Figure 15.1 include social desirability, impression management, suboptimal effort, over-reporting, and secondary gain, all of which have been described by Rogers (2008) as lacking in specificity. Arguably, defensiveness encapsulates both social desirability and impression management, while the other three terms are more specifically characterized using malingering, factitious disorder, feigning, or defensiveness.

Biasing factors influencing psychologist's interpretation of test scores

In addition to threats to score validity during the data collection process, threats to score validity exist *after* the data have been collected (see Figure 15.2). Evaluators can be influenced, without their awareness, by positive and negative qualities of the examinee.

For example, a well-groomed, polite, higher-status examinee (e.g., military officer) might be perceived by the evaluator in a more positive light, positively skewing data interpretation. This

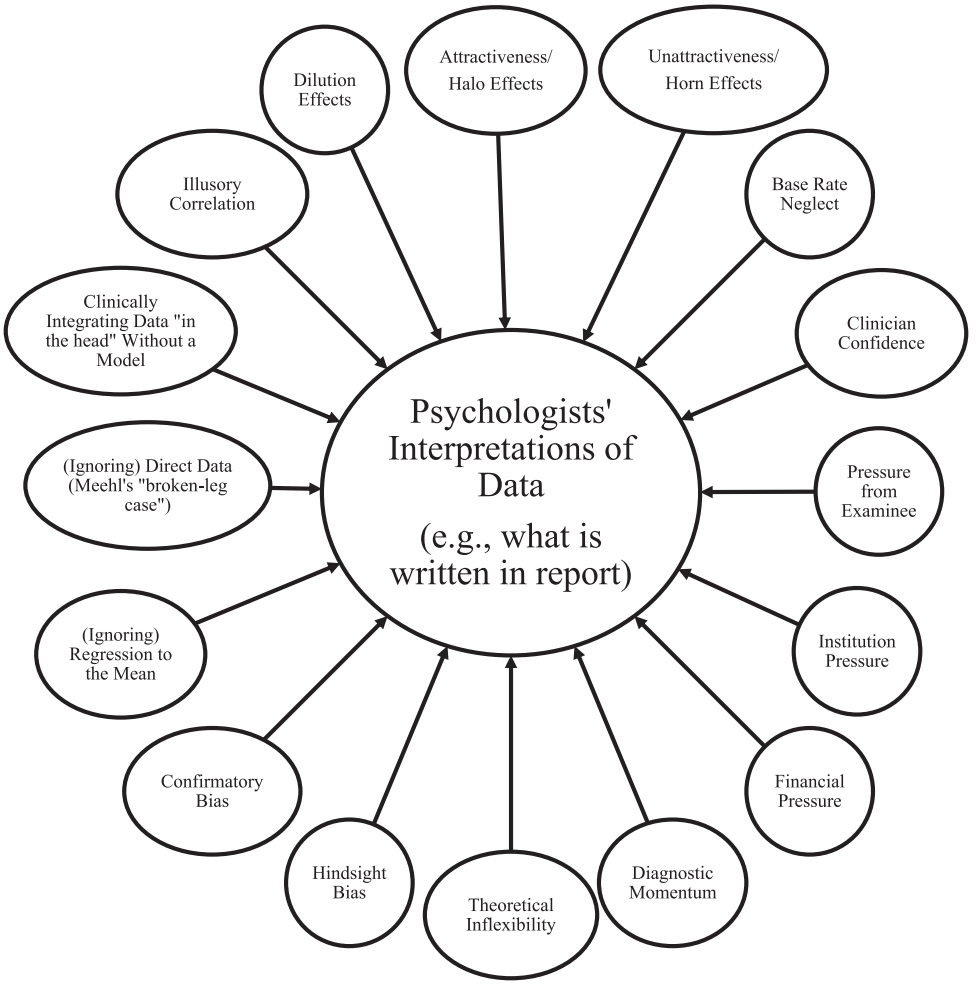

Figure 15.2 Biasing factors influencing psychologists' interpretations of data.

is the *attractiveness/halo effect*. In contrast, evaluators may develop subtle biases against examinees with unattractive characteristics, which is the *unattractiveness/horn effect*. In both scenarios, factors of tangential importance (i.e., examinee attractiveness) create bias in data interpretation.

A similar biasing effect can occur when other factors of tangential importance dilute the impact of valid variables during data interpretation. As an example, Whitehead (1986) reported that diagnostic decision-making accuracy declined among clinicians given Rorschach scores. In other words, diagnostic accuracy was better *without* the Rorschach scores. To further appreciate the *dilution effect*, consider the following from Faust, Bridges, and Ahern (2009):

> A ceiling on predictive accuracy is commonly approached or reached once the three to five most valid, and least redundant, variables have been identified. Beyond this point, additional information tends to increase confidence, not accuracy. More problematically, at times continuing to add information may decrease accuracy because weaker predictors may dilute the stronger predictors. (p. 14)

As the foregoing quote asserts, clinician over-confidence can erode accuracy, because confidence and accuracy do not consistently correlate: "confidence accounts for 2% of variance in judgment accuracy" (Miller, Spengler, & Spengler, 2015; p. 9). Hence, clinicians are encouraged to proceed humbly when interpreting data, to minimize the likelihood that overconfidence will skew data interpretation, thereby resulting in less valid conclusions.

When interpreting psychological and neuropsychological data, it is essential to consider the *base rate* of occurrence for the event or phenomenon because extremely uncommon events are harder to correctly identify, while extremely common events are easier to correctly identify. With specific respect to validity assessment, two examples are worth describing. First, it is known that examinees with healthy nervous systems score low on neuropsychological tests at a certain base rate (Binder, Iverson, & Brooks, 2009). When this rate is not considered, evaluators are more likely to over-pathologize. Second, failure of validity tests also occurs at a rate, although, as described in this chapter, that rate varies by population. When clinicians apply the wrong failure rate to examinees who do not "pass" validity tests, they are likely to make incorrect conclusions about score validity. Rogers and Bender (2013) called this the "fall-through-the-ice mentality" (p. 529). Like a hapless ice skater who drowns after falling through the surface of a frozen pond, evaluators cease any further, or nuanced, explanation of validity data after failure of a single validity test, and the examinee is immediately labeled with a pejorative term (e.g., malingerer, dissimulator), ending any further consideration of the validity data.

Base rates are also important when interpreting symptoms that are suggestive of a disorder. Just as clinicians need to consider how often non-deceivers fail validity tests, they need to consider how often non-disordered members of a population present with symptoms suggestive of a disorder. Failure to do this can result in the clinician buying into what is called an *illusory correlation* (Borum, Otto, & Golding, 1993; Faust & Ahern, 2012).

Psychological and neuropsychological tests are not perfectly reliable, meaning that different scores can be obtained under different circumstances. The more unreliable a test, the more vulnerable test scores are to what is called *regression to the mean* (Campbell & Kenny, 1999), which is when extreme scores do not replicate upon subsequent testing. Instead, scores regress back toward the most likely score, which is the mean. This happens when individuals get an unusually high, or low, score. Therefore, clinicians should consider the likelihood that an extreme score is artifactual, rather than accurate, and that it may not replicate upon future testing.

Validity assessment boils down to decision-making precision by the clinician, which can be carried out "in the head" of the clinician, using clinical/professional judgment, or using more

structured methods, including actuarial models, which are mathematical formulas developed through research to optimize decision-making accuracy. If the question of interest can be answered with an actuarial model, it is then the preferred decision-making method, given the vast literature on the superiority of actuarial methods, compared with clinical/professional judgment (Dawes, Faust, & Meehl, 1989; Grove, 2005; Grove & Meehl, 1996; Meehl, 1954). In reality, however, actuarial models are not available for every important decision, including whether a particular examinee, in a particular set of circumstances, has produced valid psychological or neuropsychological data. Not having an actuarial model to judge score validity puts clinicians in a difficult position, because they must render a decision about the validity of an examinee's behavior without the help of a mathematical model that does all the hard thinking. Therefore, it is important for clinicians making this judgment call to acknowledge the limitations of their "in-the-head" decision-making abilities, especially since researchers have shown that clinical judgment about the validity of psychological and neuropsychological data is questionable (Faust, Hart, Guilmette, & Arkes, 1988; Heaton, Smith, Lehman, & Vogt, 1978; Rosenhan, 1973; Trueblood & Binder, 1997; Walters, White, & Greene, 1988).

Rigid, inflexible adherence to decision rules can be just as problematic as confidently making clinical judgments about the validity of psychological and neuropsychological data "in the head." Perhaps the most compelling example of this was provided by the strongest advocate for actuarial models, Paul Meehl (1954), who described a hypothetical example involving an actuarial model that could predict, with a 0.90 probability, that a professor would go to the movies on a particular night. As the example goes, the professor breaks his leg, throwing off the accuracy of the actuarial model. Because the professor breaks his leg (a rare and unexpected circumstance *not* built into the actuarial model), the 0.90 accuracy rate no long holds. The important takeaway from Meehl's "broken-leg" case is that clinical judgment is still important. Hence, choosing between actuarial models and clinical judgment is, perhaps, a false dichotomy. Both methods for integrating psychological and neuropsychological data have a place.

While not an actuarial model, per se, Slick, Sherman, and Iverson (1999) developed what is now called the "Slick criteria" for Malingered Neurocognitive Dysfunction (MND). Originally, these authors proposed three categories of MND: (1) *possible* MND (i.e., examinee has an external incentive to deceive and there is evidence of discrepancy in their self-report or they fail a symptom validity test); (2) *probable* MND (i.e., examinee has an external incentive to deceive and there is stronger evidence of deception in the data [e.g., two or more failed forced-choice validity tests, but *not* below chance-level scoring]); and (3) *definite* MND (i.e., incentive to deceive and below chance-level scoring on forced-choice validity testing). Slick and Sherman (2013) revised and extended their criteria to include definite and probable *Primary* MND, *Secondary* MND (i.e., when MND criteria are met among those with legitimate neurodevelopmental, psychiatric, or neurological disorders of moderate or severe severity), and MND *by Proxy* (i.e., when examinee deception is influenced by another individual).

Presently, attempts are underway to operationally define deception in a manner that falls short of an actuarial model but which goes beyond "in-the-head" clinical judgment. What is *not* apparent, however, is the extent to which this more structured approach improves decision-making about the validity of psychological and neuropsychological data, especially in all populations and circumstances. Moreover, structured approaches like the "Slick criteria" may fuel unjustified confidence about appraising validity (Trueblood & Binder, 1997), a liability that is not specific to mental health professions (Curci, Lanciano, Battista, Guaragno, & Ribatti, 2019; Gillis & Moran, 1981; Meyer, Payne, Meeks, Rao, & Singh, 2013).

Another category of biasing influence, operating after psychological and neuropsychological data have been collected, involves outside pressure, from the examinee (or their supporters),

from the Department of Veterans Affairs (VA), or from another party paying the evaluator. Evaluators commonly feel pressure to make clinical decisions that examinees will embrace, including *not* calling them out as deceptive. Meanwhile, there can be pressure on contracted evaluators *not* to question the motives and intentions of military personnel and veterans. Poyner (2010), for example, effectively articulated her disappointing experience as an independently hired psychologist "to conduct compensation and pension evaluations with military veterans seeking VA disability benefits for PTSD and/or Traumatic Brain Injury (TBI)" (p. 130). Despite recommendations to assess for deception with veterans, Dr. Poyner was released from her VA contract because she "used malingering instruments" in her assessments.

Symptom, performance, and response validity

Larrabee (2012) published an important article distinguishing between *symptom validity tests* (SVTs) and *performance validity tests* (PVTs) and underscored the need to use these specific terms, based on "strong diagnostic discrimination for PVTs and SVTs, with a particular emphasis on minimizing false positive errors" (p. 1). Symptom validity is defined as the validity of self-reported symptoms, while performance validity is the validity of performances on tests that require the examinee to mentally exert themselves (e.g., IQ, neuropsychological, and academic achievement tests). Before this publication, it was common to refer to all validity testing, both symptom and performance, as SVTs and/or to use various other terms with less definitional clarity (e.g., "poor or inadequate effort" or "biased responding"). Of note, Larrabee's SVT/PVT terminology is not mutually exclusive with the Rogers and Bender (2013) terminology described earlier.

In 2014, Bush (2014) proposed an additional term, *response validity*, which describes "the accuracy of information provided by examinees during clinical interviews" (p. 434). While little has been written about this newer validity term, it can be operationalized by comparing historical self-report to other data points with potentially more validity. Among soldiers, response validity can be of particular relevance with regard to historical substance use, both prescribed and recreational substance use, given the motivation to underreport problematic substance-use patterns (Harrison, 1997). Hence, comparing self-report to pharmacy and legal (e.g., DWI) records can help identify discrepancies.

Many of the biasing factors in Figure 15.1 can be ruled out, or in, using SVTs and PVTs. Defensiveness, for example, can be measured using validity scales from questionnaires like the Personality Assessment Inventory (PAI) (Morey, 2007) and Minnesota Multiphasic Personality Inventory-2, Restructured Form (MMPI-2-RF) (Ben-Porath, 2012, 2013; Ben, Porath, & Tellegen, 2008). The same holds for intentional over-reporting of symptoms (i.e., dissimulation, feigning, malingering, and factitious disorder). However, it is important to emphasize that these SVTs are not "malingering" tests, per se, since malingering is just one possible explanation for SVT failure.

Reading skill, boredom, fatigue, and uncooperativeness can also affect how examinees report symptoms, which can be detected using SVTs. For example, the Inconsistency (ICN) scale on the PAI "reflects the consistency with which the respondent completed items with similar content" (p. 27). Likewise, the Infrequency (INF) scale on the PAI "is useful in identification of people who complete the PAI in an atypical way because of carelessness, confusion, reading difficulties, or other sources of random responding" (p. 28). Failure on SVTs can be a result of deception (e.g., malingering), but also reading skill, boredom, fatigue, and so forth. Therefore, clinicians must critically think through SVT failures. That said, SVTs can be invaluable in the assessment of valid self-reporting.

Like SVTs, PVTs are vital in the assessment of valid performances on tests involving mental effort. Accordingly, PVT results can help detect deception (e.g., malingering) involving low scores on cognitive tests.

Misconceptions about validity assessment

Although not an exhaustive list, the following five misconceptions about validity assessment are worth fleshing out: (1) examinees who intentionally deceive are "bad"; (2) the base rate for examinee deception is definitively known; (3) examinee deception is of trivial significance; (4) examinee deception occurs in a straightforward, predictable manner (e.g., within the same evaluation and/or across settings); and (5) cutoff scores on validity tests are extremely accurate and work equally well with all populations. Each of these misconceptions will be unpacked sequentially.

Stripped to its core, a mental health assessment is an attempt to identify and explain problematic, if not also embarrassing, behavior, such as excessive anxiety, inattention, forgetfulness, substance-use problems, depression, impulsivity, and so forth. Viewed this way, it is understandable why examinees would be motivated to hide their problematic behavior. It is also understandable why examinees would use deception in an attempt to "get" something, such as money for having a disability or justification for avoiding a difficult, or life-threatening, experience (e.g., active combat). When examinee deception is put on a finer scale of measurement, it becomes easier to appreciate that it is not a dichotomous, "good versus bad" phenomenon. This continuum of invalid responding has special relevance to the military, where malingerers are not just scorned, but court-martialed. In short, automatically concluding that examinee deception equates to "bad" behavior oversimplifies an important issue, which precludes an opportunity to understand the examinee on a more nuanced level.

As discussed earlier in relation to examinees with healthy nervous systems scoring low on neuropsychological tests at a known rate (Binder et al., 2009), familiarity with the base rate of occurrence of an important behavior is crucial because accurate measurement of the important behavior hinges on how often it is likely to be encountered in a given population. Unfortunately, it is not uncommon for mental health evaluators to both under- and overestimate the base rate of deception. The former mistake leads to false-negative conclusions about deception, whereas the latter mistake causes false-positive conclusions about deception.

The obvious question, then, pertains to the base rate for deception. Unfortunately, there is no easy answer to this question because the base rate for deception seems to vary by population and circumstance, and accurate estimates of deception have not been established. Beyond divergent *military-specific* deception base rate estimates (AFHSC, 2013; Armistead-Jehle, 2010; Denning & Shura, 2017; Freeman et al., 2008; Lande & Williams, 2013; Wall et al., 2013; Whitney et al., 2009), the following *general* (i.e., not military-specific) estimates of deception also vary dramatically: 15% (Rogers, Harrell, & Liff, 1993), 15% to 48% (Youngjohn, Burrows, & Erdal, 1995), 38.5% (Mittenberg, Patton, Canyock, & Condit, 2002), 40% (Larrabee, 2003), 1% to 20% (median values) (Sharland & Gfeller, 2007), 5% to 30% (Slick, Tan, Strauss, & Hultsch, 2004), 16% (Rogers, Sewell, & Goldstein, 1994), 15% to 17% (Frederick, 2000), 3.4% (Ruff, Klopfer, & Blank, 2016), and 0% to 30% (Young, 2015).

Clinicians embracing the responsibility of judging the motives for others' problematic behavior leads to awareness about what happens when examinee deception is not identified and when examinee deception is incorrectly identified. With the first scenario, failure to detect deception, examinees are likely to receive disability-based support that is extremely expensive or to avoid unpleasant duty when there is no justifiable medical or psychological reason. For

example, Frueh, Grubaugh, Elhai, and Buckley (2007) reported a 79.5% increase in payments to veterans with PTSD between 1999 and 2004, in contrast to a 12.2% increase in payments made to veterans with other disabilities during this same time period. Viewed through a financial lens, PTSD-related payments to veterans from 1999 to 2004 increased by 148.8% (i.e., $4.3 billion each year), compared to an increase of 41.7% in payments for other disabilities. In the second set of circumstances, incorrectly identifying deception, great harm can come to the examinee, including severe legal consequences under Article 115 of the UCMJ. Therefore, the consequential significance of accurate, and inaccurate, decision-making about examinee deception cannot be overstated; it is huge.

Deception occurs in unpredictable ways, sometimes within the same evaluation, but also across different settings (Rogers, Salekin, Sewell, Goldstein, & Leonard, 1998). Accordingly, a clinician might encounter powerful evidence of deception earlier in an assessment process, before the examinee has developed a feel for the demands of the circumstances, but later find less, or no, evidence of deception. Likewise, deception may occur as a function of other factors, such as the quality of the informed consent process (e.g., forewarning examinees about the importance of full participation and engagement, versus failing to do so) or the referral question (Rogers & Salekin, 1998). Hence, mental health evaluators should operate with humility about their ability to correctly identify deception.

Rogers and Bender (2013) emphasized the lack of "laser accuracy of cut scores" (p. 519) on tests of validity, meaning that there is a margin of error associated with any given validity test cut score. Accordingly, they argued that rigid application of a cut score to a given examinee can result in spurious findings, the "fall-through-the-ice mentality" described previously. To minimize this mistake, Rogers and Gillard (2011) recommend using the standard error of measurement (SEM) associated with the test to minimize error. Namely, they assert that validity scores within 1 SEM should be considered *indeterminate*, rather than making more definitive conclusions about the presence or absence of deception. Additionally, the cutoff scores reported in test manuals are sometimes found to be less accurate than scores later established through more extensive research after publication of the test. For this reason, clinicians should remain abreast of ongoing research on validity assessment measures.

Another strategy to avoid false-positive conclusions based on a single failed validity test is to establish a threshold for score invalidity based on *multiple* failed validity tests, not just one. Schutte and Axelrod (2013) described the benefits and liabilities with this approach, including the increased potential for false-positive conclusions when validity tests are strongly correlated (i.e., are multicollinear). In her discussion of neurocognitive symptom validity, Boone (2013b) applied this multiple-cut-score approach to those with dementia:

> [I]f a patient fails 85% of the [PVTs] administered, this would be comparable to the performance of patients with advanced dementia … and such a test taker would not be able to live independently, drive, handle finances, or manage other important ADLs. Thus, if a patient failing 85% of [PVTs] does not demonstrate markedly impaired function in ADLs, the [PVT] failure can be relied upon as evidence of noncredible neurocognitive performance. (p. 47)

Two things should be evident with regard to validity test cut-scores: (1) a rigid, or cavalier, application of cut-scores is likely to increase decision-making error, especially when assessing those with legitimate neurocognitive and/or intellectual deficits (Graue, 2006; Shandera et al., 2010), and, as emphasized previously, (2) evaluators should not use a one-size-fits-all approach to cut-score use.

Conclusions

Psychological assessment in the United States flourished as a result of World War I because it was necessary to accurately identify those who should be excluded from military service, as well as those of uncommon aptitude. While aptitude testing remains important, mental health professionals working with military personnel must also accurately characterize mental health problems. To effectively accomplish these tasks, evaluators should endeavor to increase the quality of data *collection* and *interpretation*. A vital step in data collection is to properly utilize SVTs and PVTs. It is equally important to stay aware of limitations imposed by "in-the-head" clinical judgment about the presence of deception, as well as misconceptions about validity assessment. While actuarial models are not available for every challenging decision faced by the clinician, other methods (e.g., the revised and extended "Slick criteria" for MND) offer clinicians a more structured way of identifying data among military personnel that are of questionable validity due to examinee deception. Thus, a multi-method approach to validity assessment is ideal. Because accurate identification and understanding of invalid psychological and neuropsychological data among military personnel remains a challenging and nuanced process, it should be carried out thoughtfully and with an awareness of available research, which continues to evolve.

References

Armed Forces Health Surveillance Center. (2013). Malingering and factitious disorders and illnesses, active component, US Armed Forces, 1998–2012). *Medical Surveillance Monthly Report*, *20*(7), 20–24.

Armistead-Jehle, P. (2010). Symptom validity test performance in US veterans referred for evaluation of mild TBI. *Applied Neuropsychology*, *17*(1), 52–59.

Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis, MN: University of Minnesota Press.

Ben-Porath, Y. S. (2013). *Using the MMPI-2-RF validity scales in forensic assessments*. Retrieved from http://www.apadivisions.org/division-18/publications/newsletters/gavel/2013/07/forensic-assessments.aspx

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Manual for Administration, Scoring, and Interpretation*. Minneapolis: University of Minnesota Press.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*(1), 31–46.

Boone, K. B. (2013a). *Clinical Practice of Forensic Neuropsychology: An Evidence-Based Approach*. New York, NY: Guilford Press.

Boone, K. B. (2013b). A reconsideration of the Slick et al. 1999 criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of Feigned Cognitive Impairment* (pp. 29–49). New York, NY: The Guilford Press.

Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *The Journal of Psychiatry & Law*, *21*(1), 35–76.

Bush, S. S. (2014). Assessment of symptom and performance validity in veterans. In S. S. Bush (Ed.), *Psychological Assessment of Veterans* (pp. 434–452). New York, NY: Oxford University Press.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., …. Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426.

Campbell, D. T., & Kenny, D. A. (1999). *A Primer on Regression Artifacts*. New York, NY: Guilford Publications.

Curci, A., Lanciano, T., Battista, F., Guaragno, S., & Ribatti, R. M. (2019). Accuracy, confidence, and experiential criteria for lie detection through a videotaped interview. *Frontiers in Psychiatry*, *9*, 1–14.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674.

Denning, J. H., & Shura, R. D. (2017). Cost of malingering mild traumatic brain injury-related cognitive deficits during compensation and pension evaluations in the veterans benefits administration. *Applied Neuropsychology: Adult*, *26*(1), 1–16.

Faust, D., & Ahern, D. C. (2012). Clinical judgment and prediction. In D. Faust (Ed.), *Coping with Psychiatric and Psychological Testimony* (6 ed., pp. 147–208). New York, NY: Oxford University Press.

Faust, D., Bridges, A. J., & Ahern, D. C. (2009). Methods for the identification of sexually abused children: Issues and needed features for abuse indicators. In K. Kuehnle & M. Connell (Eds.), *The Evaluation of Child Sexual Abuse Allegations: A Comprehensive Guide to Assessment and Testimony* (pp. 3–19). New York: Wiley.

Faust, D., Hart, K. J., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, *19*(5), 508–515.

Frederick, R. I. (2000). Mixed group validation: A method to address the limitations of criterion group validation in research on malingering detection. *Behavioral Sciences & the Law*, *18*(6), 693–718.

Frederick, R. I. (2012). Malingering/cooperation/effort. In D. Faust (Ed.), *Coping with Psychiatric and Psychological Testimony* (6 ed., pp. 229–247). New York, NY: Oxford University Press.

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research*, *158*(3), 374–380.

Frueh, B. C., Grubaugh, A. L., Elhai, J. D., & Buckley, T. C. (2007). US Department of Veterans Affairs disability policies for posttraumatic stress disorder: Administrative trends and implications for treatment, rehabilitation, and research. *American Journal of Public Health*, *97*(12), 2143–2145.

Gillis, J. S., & Moran, T. J. (1981). An analysis of drug decisions in a state psychiatric hospital. *Journal of Clinical Psychology*, *37*(1), 32–42.

Graue, L. O. (2006). *Detection of malingered mental retardation*.

Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*, *61*(10), 1233–1243.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293–323.

Harrison, L. (1997). The validity of self-reported drug use in survey research: An overview and critique of research methods. *National Institute on Drug Abuse Research Monograph*, *167*, 17–36.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*(5), 892–900.

Lande, R. G., & Williams, L. B. (2013). Prevalence and characteristics of military malingering. *Military Medicine*, *178*(1), 50–54.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*(3), 410–425.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(04), 625–630.

Meehl, P. E. (1954). *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press.

Meyer, A. N., Payne, V. L., Meeks, D. W., Rao, R., & Singh, H. (2013). Physicians' diagnostic accuracy, confidence, and resource requests: A vignette study. *JAMA Internal Medicine*, *173*(21), 1952–1958.

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of confidence and judgment accuracy in clinical decision making. *Journal of Counseling Psychology*, *62*(4), 553–567.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1094–1102.

Morey, L. C. (2007). *Personality Assessment Inventory Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology*, *47*(2), 409–410.

Peterson, E., & Peterson, R. (2015). Understanding deception from a developmental perspective. In M. Kirkwood (Ed.), *Validity Testing in Child and Adolescent Assessment: Evaluating Exaggeration, Feigning, and Noncredible Effort* (pp. 42–61). New York, NY: The Guilford Press.

Poyner, G. (2010). Psychological evaluations of veterans claiming PTSD disability with the Department of Veterans Affairs: A clinician's viewpoint. *Psychological Injury and Law*, *3*(2), 130–132.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie*, *28*, 286–340.

Rogers, R. (2008). An introduction to response styles. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (3 ed., pp. 3–13). New York, NY: Guilford Press.

Rogers, R., & Bender, S. D. (2013). Evaluation of malingering and related response styles. In R. K. Otto (Ed.), *Handbook of Psychology* (2 ed., pp. 517–540). Hoboken, NJ: John Wiley & Sons, Inc.

Rogers, R., & Gillard, N. D. (2011). Research methods for the assessment of malingering. In B. Rosenfeld & S. Penrod (Eds.), *Research Methods in Forensic Psychology* (pp. 174–188). Hoboken, NJ: John Wiley & Sons Inc.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, *13*(3), 255–274.

Rogers, R., & Salekin, R. T. (1998). Beguiled by Bayes: A re-analysis of Mossman and Hart's estimates of malingering. *Behavioral Sciences & the Law*, *16*(1), 147–153.

Rogers, R., Salekin, R. T., Sewell, K. W., Goldstein, A., & Leonard, K. (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., & Goldstein, A. M. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior*, *18*(5), 543–552.

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, *179*(4070), 250–258.

Ruff, R. M., Klopfer, J., & Blank, J. (2016). Homogenous base rates for malingering in neuropsychological examination of litigants. *Psychological Injury and Law*, *9*(2), 143–153.

Schutte, C., & Axelrod, B. N. (2013). Use of embedded cognitive symptom validity measures in mild traumatic brain injury cases. In D. A. Carone & S. S. Bush (Eds.), *Mild Traumatic Brain Injury: Symptom Validity Assessment and Malingering* (pp. 159–181). New York, NY: Springer Publishing Company.

Shandera, A. L., Berry, D. T., Clark, J. A., Schipper, L. J., Graue, L. O., & Harp, J. P. (2010). Detection of malingered mental retardation. *Psychological Assessment*, *22*(1), 50–56.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology*, *22*(2), 213–223.

Slick, D. J., & Sherman, E. M. (2013). Differential diagnosis of malingering. In D. A. Carone & S. S. Bush (Eds), *Mild Traumatic Brain Injury: Symptom Validity Assessment and Malingering* (pp. 57–72). New York, NY: Springer Publishing Company.

Slick, D. J., Sherman, E. M., & 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. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology*, *19*(4), 465–473.

Stone, D. C., & Boone, K. B. (2007). Feigning of physical, psychiatric, and cognitive symptoms: Examples from history, the arts, and animal behavior. In K. B. Boone (Eds.), *Assessment of Feigned Cognitive Impairment* (pp. 3–12). New York, NY: Guilford Press.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests*. Administration, Norms, and Commentary. New York: Oxford University Press.

Trueblood, W., & Binder, L. M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology*, *12*(1), 13–27.

U.S. v. Greer, No. 96–11443, 158 F.3d 228 (Court of Appeals, 5th Circuit 1998). Retrieved from https://casetext.com/case/us-v-greer-18

Vore, D. (2007). The disability psychological independent medical evaluation: Case law, ethical issues, and procedures. In A. M. Goldstein (Ed.), *Forensic Psychology: Emerging Topics and Expanding Roles* (pp. 489–510). Hoboken, NJ: John Wiley & Sons, Inc.

Wall, A. M., Graver, C. J., & Shurak, N. R. (2013). Base rates of suboptimal effort in a broad clinical population. *The Clinical Neuropsychologist*, *27*(4), 607–608.

Walters, G. D., White, T. W., & Greene, R. L. (1988). Use of the MMPI to identify malingering and exaggeration of psychiatric symptomatology in male prison inmates. *Journal of Consulting and Clinical Psychology*, *56*(1), 111–117.

Whitehead, W. C. (1986). *Clinical decision making on the basis of Rorschach, MMPI, and automated MMPI report data*. University of Texas Health Science Center at Dallas, *Unpublished doctoral dissertation*.

Whitney, K. A., Shepard, P. H., Williams, A. L., Davis, J. J., & Adams, K. M. (2009). The medical symptom validity test in the evaluation of operation Iraqi freedom/operation enduring freedom soldiers: A preliminary study. *Archives of Clinical Neuropsychology*, *24*(2), 145–152.

Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. New York, NY: Henry Holt & Company.

Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law*, *8*(3), 188–199.

Youngjohn, J. R., Burrows, L., & Erdal, K. (1995). Brain damage or compensation neurosis? The controversial post-concussion syndrome. *The Clinical Neuropsychologist*, *9*(2), 112–123.