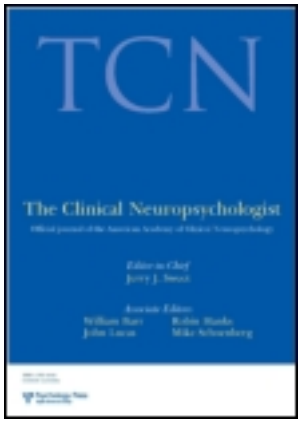

The Clinical Neuropsychologist
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ntcn20

American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering

Robert L. Heilbronner, Jerry J. Sweet, Joel E. Morgan, Glenn J. Larrabee, Scott R. Millis & Conference Participants

Available online: 12 May 2010

To cite this article: Robert L. Heilbronner, Jerry J. Sweet, Joel E. Morgan, Glenn J. Larrabee, Scott R. Millis & Conference Participants (2009): American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering, The Clinical Neuropsychologist, 23:7, 1093-1129

To link to this article: http://dx.doi.org/10.1080/13854040903155063

The Clinical Neuropsychologist, 23: 1093–1129, 2009
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854040903155063

Psychology Press
Taylor & Francis Group

CE

AMERICAN ACADEMY OF CLINICAL NEUROPSYCHOLOGY CONSENSUS CONFERENCE STATEMENT ON THE NEUROPSYCHOLOGICAL ASSESSMENT OF EFFORT, RESPONSE BIAS, AND MALINGERING

Robert L. Heilbronner, Jerry J. Sweet, Joel E. Morgan, Glenn J. Larrabee, Scott R. Millis, and Conference Participants[1]

During the past two decades clinical and research efforts have led to increasingly sophisticated and effective methods and instruments designed to detect exaggeration or fabrication of neuropsychological dysfunction, as well as somatic and psychological symptom complaints. A vast literature based on relevant research has emerged and substantial portions of professional meetings attended by clinical neuropsychologists have addressed topics related to malingering (Sweet, King, Malina, Bergman, & Simmons, 2002). Yet, despite these extensive activities, understanding the need for methods of detecting problematic effort and response bias and addressing the presence or absence of malingering has proven challenging for practitioners. A consensus conference, comprised of national and international experts in clinical neuropsychology, was held at the 2008 Annual Meeting of the American Academy of Clinical Neuropsychology (AACN) for the purposes of refinement of critical issues in this area. This consensus statement documents the current state of knowledge and recommendations of expert clinical neuropsychologists and is intended to assist clinicians and researchers with regard to the neuropsychological assessment of effort, response bias, and malingering.

[1]Conference Participants (by working group): *Definitions and Differential Diagnosis*: Robert L. Heilbronner (Chair), Kevin J. Bianchini, Paul Kaufmann, Daniel J. Slick, H. Gerry Taylor. *Ability Issues*: Jerry J. Sweet (Chair), Kyle Brauer Boone, Shane S. Bush, Kevin W. Greve, Thomas J. Guilmette. *Somatic Issues*: Glenn J. Larrabee (Chair), Manfred F. Greiffenstein, Nathaniel W. Nelson, Julie Suhr, David T. R. Berry*. *Psychological Issues*: Joel E. Morgan (Chair), Robert L. Denney, Robert J. McCaffrey, Christopher L. Grote, Roger O. Gervais. *Research Evidence and Scientific Issues*: Scott R. Millis (Chair), William B. Barr, Jacobus Donders, Grant L. Iverson, Martin L. Rohling. *External Review Panel*: Laurence M. Binder, Paul Lees-Haley, George J. Demakis, Wiley Mittenberg, Richard I. Frederick.

*Dr. Berry was unable to attend the meeting, but reviewed and approved the draft work of the working group.

Conference participants were not compensated for their participation or reimbursed for expenses incurred as a result of participation. The authors acknowledge the assistance of Dawn Giuffre Meyer in the preparation of this manuscript.

Address correspondence to: Jerry J. Sweet. E-mail: jerrysweet@uchicago.edu
Accepted for publication: June 30, 2009.

DEVELOPMENT OF THE CONSENSUS CONFERENCE AND THE CONSENSUS STATEMENT

Organizers Robert Heilbronner and Jerry Sweet developed the idea for the Consensus Conference, with the need for the conference based on the degree of interest expressed on the part of forensic neuropsychology experts, clinical researchers, and both junior and senior colleagues in the field. It was apparent from high attendance at relevant continuing education presentations during the past decade that a substantial proportion of practicing neuropsychologists were regularly seeking guidance based on relevant research and the advice of identified expert colleagues. The desired guidance appeared to focus primarily on the detection of problematic effort and other forms of response bias, such as biased symptom reporting, related to diagnosing or ruling out neuropsychological malingering. The American Academy of Clinical Neuropsychology (AACN) Board of Directors approved sponsorship of the proposed conference in June 2007, entrusting Heilbronner and Sweet with the organization of the conference, to be held at the annual meeting of the Academy in Boston in June 2008, with the resulting statement to be published in the official journal of the Academy, *The Clinical Neuropsychologist*.

The primary goal of the consensus conference was to produce a consensus statement, developed by prominent expert neuropsychologists, based on sound practice principles and reflecting the broad scientific literature relevant to the measurement of effort, response bias, and neuropsychological malingering. There is a common misperception that clinical neuropsychologists often find points of disagreement on clinical and forensic topics. In actuality all science-driven healthcare specialties create progress by a process of challenging current and new ideas through intellectual discourse and empirical hypothesis testing. In this manner, clinical neuropsychology is no different. For this reason it was expected that clinical and forensic experts who are involved regularly in presenting at professional meetings and/or who frequently publish empirical peer-reviewed research on the subject matter of the conference would in fact be able to arrive at points of consensus. In turn, these points of consensus can aid neuropsychology practitioners and researchers.

The organizers identified five relevant content areas to be addressed at the consensus conference and reviewed a lengthy bibliography (Sweet, 2009) to identify broad and specific topic experts for each. Relevant to each content area, literature citations and professional conference programs were reviewed to identify relevant experts. Of the eventual list of 25 experts identified, five individuals (Heilbronner, Larrabee, Millis, Morgan, Sweet) were identified as content section working group chairs. These chairs met in February 2008 to discuss the proposed structure and process of the conference. During this initial meeting, a number of consensus statements published previously by other professional organizations were reviewed.

The working group chairs agreed in advance to identify, with the assistance of the members of their working groups, pre-conference reading materials. The list of each group's pre-conference readings (together with the agenda of the conference) was circulated to all 25 invitees before the meeting, and can be found in the Appendix.

PROCESS OF CREATING CONSENSUS AND WRITING THE CONSENSUS STATEMENT

The participants of the conference met on June 17, 2008, the day prior to the AACN Annual Conference and Workshops in Boston. The conference began with a review of the general approach and purpose of a consensus conference.

Conflict of interest forms were completed at the beginning of the conference by all participants because of the possibility that specific commercially available tests authored by and accruing royalties to one or more participants might be mentioned in the eventual consensus statement. Comparison of potential conflicts of interest with the final draft of the statement did not reveal conflict of interest for any participant. More generally, the salience and relevance of active involvement in forensic activities was a qualification for invitation to be part of the consensus conference, either as a participant or as an external reviewer. The very nature of a consensus conference, which requires agreement among all contributors, effectively negates the influence of any single participant's or external reviewer's personal preference or bias in forensic practice.

The conference alternated between working group breakouts and overall group discussions. Each working group had been provided an outline in advance, to assist in the completion of a work product for each group. From this outline, each group began to develop its portion of the consensus statement. At multiple points in the initial full day of the conference, working group chairs summarized points of consensus to the entire group of participants to assess the larger group's viewpoints and possible consensus. After the end of the first day, working group chairs constructed a PowerPoint presentation, which was presented to a large audience of general AACN conference attendees in the form of a 3-hour workshop on June 19. Each of the working group chairs presented points of consensus relevant to their content areas, for discussion with and input from the audience members. The conference participants reconvened on the afternoon of June 20, following the close of the AACN meeting, to discuss and integrate the input from the workshop attendees. Subsequently, working groups worked via phone and e-mail to construct the five sections of this statement, which the organizers integrated into one document. The draft statement then underwent additional review by all the conference participants, as well as a five-member independent panel of external reviewers.[2] Thus, the content of this Consensus Conference Statement reflects the input of 30 clinical neuropsychologists with expertise relevant to the neuropsychological assessment of effort, response bias, and malingering.

DEFINITIONS AND DIFFERENTIAL DIAGNOSIS

As noted in the introduction to this Consensus Conference Statement, the majority of research and clinical work on the neuropsychological assessment of effort, response bias, and malingering has occurred during the last 20 years. During this time various descriptors and terms have been applied to convey salient relevant

[2] External Review Panel: Laurence M. Binder, Paul Lees-Haley, George J. Demakis, Wiley Mittenberg, Richard I. Frederick.

concepts, sometimes in a manner that is helpful, and other times less so. We aspire in this section to clarify terminology, in a manner that facilitates an understanding of relevant concepts for clinical practice and scientific investigation.

In the overall process of conducting an assessment, when there is concern about insufficient effort or malingering, a distinction can be made between the processes of ''detection'' and ''diagnosis.''

Detection

Neuropsychologists are concerned with cognitive, emotional, and physical capacities and symptoms. When neuropsychologists use indicators developed in studies that feature malingering and non-malingering groups, the classification accuracy statistics describe the capacity of the indicator (at different score levels) to detect malingering (see more about this below in the ''Research Design'' section). The *detection* of malingering can occur in the neuropsychological assessment of any individual during the course of assessing symptoms, abilities/capacities, and functionality. There are two strategies employed by examinees that are of interest, which undermine the validity of the examination. These involve attempts to intentionally create the appearance of disability, which is a common primary goal of malingering (Bianchini, Greve, & Glynn, 2005). These strategies are to intentionally diminish or reduce capability and intentionally exaggerate symptom complaints.

Various terms have been used by researchers and clinicians to describe the behaviors of interest in the identification of intentionally exaggerated symptoms and diminished or reduced capability. In particular, when related to the validity of ability measurements, clinical researchers have chosen a number of words to convey problems with *effort*. These terms include, among others: insufficient effort, inadequate effort, and poor effort. There does not appear to be consensus on the preferred descriptor for effort, yet it can be noted in these three examples that a useful descriptor of problematic effort is one that clearly conveys a substantial negative impact that has the potential of invalidating measurement of ability. Measures used to identify problematic effort are often identified as *effort tests*, which are considered to be in a category of measures that evaluate validity of symptoms, known as *symptom validity tests* (SVTs). In this vein, it is important to clarify that these *effort tests* actually require little effort or ability, as they typically are normally performed (and in some cases, perfectly performed) by a wide range of patients who have bona fide neurologic, psychiatric, or developmental problems. Failure on such measures indicates that not enough effort was expended in the direction of capable performance. Some experts prefer to consider that failure on an effort test shows that considerable effort to perform poorly was expended. Validity indicators that are ''built in'' to standard neuropsychological measures are referred to as *embedded measures*. The focus of specialized symptom validity tests and embedded indicators is considered to be the measurement of *performance validity* or *response validity*. Most procedures designed to detect a response bias relevant to malingering are aimed at identifying a *negative response bias*. This element of the detection process dealing with invalid performances on measures of neuropsychological function will be discussed more extensively in the Ability Issues section of this Consensus Conference document.

Intentional exaggeration of symptom complaints is a separate issue in the malingering detection process. Common terms related to this examinee behavior include *symptom exaggeration* and *symptom magnification*. Indicators of symptom exaggeration are often incorporated into psychological tests that elicit self-report of symptoms. This topic will be discussed more extensively in the Somatic and Psychological Issues sections of this document.

When considering neuropsychological test performance, concerns regarding effort are frequently related to consideration of whether an examinee is malingering. However, simply equating "poor effort" with malingering is an oversimplification. For example, an examinee who is malingering may simply put forth poor effort on neuropsychological testing as a component of his or her approach to malingering. A different examinee, also malingering, may expend considerable effort to avoid detection while being examined. This is a complex conceptual issue. The process of detecting malingering is one in which consideration is given to multiple dimensions of behavior that differentiate malingering from other entities, such as factitious disorder, conversion disorder, cogniform disorder (Delis & Wetter, 2007), and somatoform disorder.

The dimension of *effort level* can be conceptualized as occurring on a continuum and can vary within and across tests and also across other types of behaviors observed during an assessment (e.g., level of effort expended to recall autobiographical information during the interview or level of effort expended in an attempt to convincingly present with episodic memory deficits during formal testing). Other related concepts (e.g., volitional vs non-volitional, internal goal vs external goal, etc.) are best considered not as present or absent, but rather as being continua denoting the relative influence of opposite factors. For example, in some instances, examinee behavior (e.g., intentionally feigning deficits) may be for the purpose of meeting internal psychological needs (e.g., factitious disorder) or toward obtaining an external, material reward (e.g., malingering). In other instances, either an internal or external goal may be pre-eminent, or both may be equally important. Note that these dimensions can be thought of as underlying very specific behaviors, such as an examinee's response to a given test item or specific interview question. These dimensions can also be used to characterize an examinee's behavior more broadly, such as during performance on a given test or test battery or during the course of an interview.

Diagnosis

In considering the diagnosis of malingering, the clinician is explicitly making a determination of intent: more specifically, a determination of intentionally exaggerated symptoms and/or intentionally diminished capability with the goal of obtaining an external reward. The committee reached a consensus that, through application of relevant psychological and neuropsychological science, clinicians *can diagnose* malingering in some examinees. Clinicians using this scientific foundation/body of knowledge can differentiate intentionally exaggerated presentations/disorders/conditions/diagnoses (e.g., malingering and factitious disorder) from unintentionally exaggerated presentations (e.g., somatoform pain disorder, cogniform disorder). To do this, the context of the evaluation and overall presentation of

the examinees, including background information, history information gathered during interview, observations, neuropsychological tests, and measures of response bias, should be considered in this process.

For the diagnosis of malingering, there are published diagnostic classification systems (e.g., Bianchini et al., 2005; Slick, Sherman, & Iverson, 1999) that better represent current neuropsychological knowledge regarding malingering indicators than the current version of the American Psychiatric Association Diagnostic and Statistical Manual, Fourth Edition-Text Revision (DSM-IV-TR; APA, 2000). Although likely to evolve and undergo refinement across time, such neuropsychological diagnostic systems offer a reliable means of operationalizing diagnostic decisions related to the determination of malingering and are consistent with appropriate clinical (*AACN Practice Guidelines for Neuropsychological Assessment and Consultation*; AACN, 2007) and forensic (*Specialty Guidelines for Forensic Psychologists*; Committee on Ethical Guidelines for Forensic Psychologists, 1991) guidelines in explicitly recommending that clinicians incorporate multiple sources of data and information. Empirically based systems are recommended when making a diagnosis, as they provide increased reliability of the classification accuracy of findings of the various validity indicators. It is recommended that clinicians be familiar with the psychological and neuropsychological literature related to the classification accuracy of validity indicators and how well the sample of a given study generalizes to the individual being examined. Within this Consensus Conference Statement, the term *diagnosis* is used in a manner consistent with use in the relevant neuropsychological literature (e.g., Bianchini et al., 2005; Slick et al., 1999). In the current version of the DSM-IV-TR (APA, 2000), malingering is assigned a *V-code* indicating that it is not a psychiatric illness and that a disease process is not implied with the designation. Although we use the term *diagnosis,* we agree with the DSM-IV-TR that malingering is *not* a mental illness or disorder. In this sense, our use of the term diagnosis refers to assigning malingering as a designation and a descriptive term to describe intentional exaggeration, and is not a description of a disease process.

Because the diagnosis of malingering involves an explicit consideration of the *purpose* of a given behavior, the committee recognizes that an important part of the diagnostic process involves a consideration of the *context* of the evaluation. In a routine clinical context the primary gain of a *patient* is relief of some form of physical or emotional symptom. Some terms have been used and misused in describing important aspects of the forensic context, including: *gain, secondary gain*, *external gain*, and *financial gain*. In order to reach a determination that intentional exaggeration is produced as a manifestation of malingering (versus factitious disorder) the clinician must determine that the patient has something to gain from being impaired. This potential is sometimes referred to as *secondary gain*, which is often mistakenly used as a synonym for malingering. Secondary gain differs from primary, typically emotional, gain, which is often the characteristic motivation for the intentional exaggeration in factitious disorders. The concept of secondary gain is well described in the DSM-IV-TR (APA, 2000) as including: financial reward, compensated time away from work, avoidance of military duty, relief from legal consequences, and obtaining medications/narcotics. It is usually not difficult to determine whether the patient has the capacity to gain from his or her symptoms or

claimed impairment. Typical contexts that involve secondary gain potential include litigation, disability claims, criminal prosecution, and worker's compensation claims. Military service and post-service involvement with the Department of Veterans Affairs can also involve secondary gain potential, as noted in a relevant task force report that includes mention of the appropriateness of evaluating symptom validity and response bias (McCrea et al., 2008). The committee recommends that the use of the term secondary gain be limited to a *description of the context within which the evaluation is taking place* and not used as a synonym for malingering.

The term *malingering* is descriptive. In some instances malingering can be adaptive. As previously discussed, the differential diagnosis of malingering must take into consideration the role of intent. The best way to assess intent is by ruling out other possible conditions (e.g., psychological, neurological, developmental) that might otherwise explain the suspicious behavioral presentation and by requiring the presence of multiple improbable performances and/or atypical symptomatic complaints. The differential diagnosis of malingering is a clinical process that: (1) requires careful analysis on the part of the examiner, (2) is based on objective criteria; (3) incorporates indicators that have established classification accuracy, and (4) combines clinical judgment with the results of scientifically validated measures in this process.

An attempt to generalize one set of behaviors (e.g., malingering behaviors) to a specific case involves making an inference. As is true for clinical decision making, when making a determination of malingering, this inference, and the data set on which it is based, including formal test results, must have a strong scientific foundation. A salient judgment concerns how well the published literature generalizes to the examinee. Staying abreast of current knowledge facilitates scientifically based inferences about the behavior of examinees, both within and outside of the test environment.

There is consensus regarding the existence of a research foundation that neuropsychologists can rely on in determining an examinee's intent to exaggerate symptoms or perform below their capabilities in testing. Scientific literature in the field of clinical neuropsychology indicates that neuropsychologists possess evidence-based methods for detecting intentional exaggeration of cognitive and emotional functioning and somatic complaints that can be used in an empirical, clinical decision-making process. Relying on this scientific body of knowledge, neuropsychologists are capable and qualified to diagnose malingering.

The determination that malingering is present often involves the application of scientific results to a forensic question. This information can be used to assist the trier-of-fact (e.g., judge, jury) in a legal decision-making process. Neuropsychologists remain mindful of the important difference between scientifically based clinical decisions and legal adjudication. Moreover, they recognize and respect the laws and customs of the jurisdiction in which they practice when describing the behavioral presentation at issue. Finally, in performing a competent differential diagnosis of malingering, the process is best served by a dispassionate consideration of all factors related to the context in which the behavior occurs. This includes consideration of the diversity of examinees, in terms of language, culture, education, and ethnic minority

status, which may not be matched by the diversity of neuropsychological examiners in these same variables (Romero et al., 2009).

ABILITY ISSUES

Conceptual and operational definitions

Misrepresentation of abilities in any neuropsychological domain of ability (memory, sensorimotor, language, etc.) through performance, or self-report regarding performance capabilities, represents response bias. Whereas many of the ability domains could be termed *cognitive*, some, such as motor function, are more accurately depicted as *abilities*. When in the presence of the potential for external gain the valence of the response bias is negative (i.e., in the direction that would increase the likelihood or magnitude of external gain), malingering needs to be considered. External gain can take the form of monetary reward, non-monetary incentive (e.g., drug seeking), avoidance of responsibility, or escape from undesirable or intolerable conditions.

Negative response bias involving ability performance is operationalized by failure to surpass the thresholds of effort tests or embedded validity indicators within ability tests and/or response bias validity scales within self-report measures. Alternatively, gross disparity between ability test performance and real-world activities could represent response bias. With regard to ability performances, these invalid presentations: (1) are not fully explained by brain dysfunction, (2) are not reasonably attributable to variables that may in some instances moderate (e.g., education, age) or may in some instances confound (e.g., fatigue, psychological conditions) performances on ability tests, and (3) are significantly worse than, or at least different in degree or pattern from, performance standards known to reflect genuine neurological disorder.

Review of records, clinical interview, and comparison of test results to "real-world" behavior can be essential in addressing the possibility of malingering. Related to interviewing, the concepts of *false history and incomplete history* are important. These represent more than normal errors of omission in providing history or inexact provision of history. Providing a false history involves misrepresenting relevant information that is salient to issues central to a forensic consultation. Providing incomplete history involves omitting important information that could prove central to consideration of the causes of, or alternative explanations for, claimed symptoms and alleged disability. In the process of evaluating an examinee's credibility as a historian, it is important to assess whether the individual is providing a reliable and valid account of his/her history.

Tests and other psychometric procedures relied on by clinicians in judging response validity must themselves have proven validity. In fields of scientific investigation, such as clinical neuropsychology and related areas, there are multiple peer review journals in which the validity of tests and procedures related to neuropsychological response bias have been and continue to be the subject of empirical scrutiny (Sweet et al., 2002). The size of the neuropsychological literature related to assessment of response bias in the measurement of abilities and the overall quality of this literature is substantial and well developed; this area

of investigation should not be viewed as "experimental" or nascent. Indeed, the relevant scientific literature is broad and varied in its depiction of the effectiveness of specific measures, with a range of classification accuracy results. Clinicians need to be aware of these studies well beyond the original research contained within test manuals or initial publications of applications, and this requires staying current with the scientific literature in this area. Clinicians need to assign weight to specific results according to the rigor of the studies and the relevance of samples studied to the clinician's case at hand. It is expected that the quality and amount of existing research will guide the clinician's choice of measures for use in individual cases, with greater weight being given to indicators that have proven validity across multiple studies.

Types of assessment methods related to evaluating response validity of abilities

Methods of evaluating *performance*

Stand-alone cognitive effort tests. These measures have been developed specifically to evaluate task performance validity. Additional testing time is required for this class of tests. There is now abundant research evidence that stand-alone cognitive effort tests are extremely useful within forensic evaluations, which have been shown to be associated with a high risk of invalid responding. Therefore, the additional testing time is warranted, and "medically necessary" (Bush et al., 2005).

Forced-choice stand-alone cognitive effort measures are those that limit the examinee to choosing one of a fixed number of responses. With regard to stand-alone cognitive effort testing, most measures offer two response choices, such that a chance performance can be determined. Numerous clinical researchers (e.g., Greve, Binder, & Bianchini, 2009; Mittenberg, Patton, Canyock, & Condit, 2002; Slick, Tan, Strauss, & Hultsch, 2004) have investigated the base rates on forced-choice measures of *significantly below* chance performance (see statistical discussion by Frederick & Speed, 2007), which has been viewed as evidence of deliberate under-performance, and when occurring within a secondary gain context, supports a conclusion of malingering. As shown in additional research (e.g., Thompson, 2002; Tombaugh, 2002), it is now well known that invalid performance can be identified using thresholds that are well above a level that is significantly below chance.

Non-forced-choice stand-alone cognitive effort measures are those that allow a range of responses and may evaluate random responding, unrealistically slow or erroneous responding, and inconsistency of response patterns when compared to performances from well-documented disorders evaluated in a non-forensic context.

Embedded indicators within ability tests. This category of measures refers to validity indicators derived from standard clinical ability tests that have shown value in identifying non-credible or disingenuous performances. There are two forms: indicators developed for this specific purpose and traditional standard scores that have been found in post-release research to be sensitive to insufficient effort. To date, the majority of embedded indicators have been developed in

post-release research. The formats of these embedded measures are broad and can include consideration of forced-choice responding.

There are too many effort measures and embedded validity indicators with proven validity to list within this statement; extensive lists are available elsewhere (cf. Sweet, 2009). Effort measures have been the subject of numerous empirical studies, which have been summarized in reviews and textbooks (e.g., Boone, 2007; Hom & Denney, 2002; Larrabee, 2007) and meta-analytic research (Vickery, Berry, Inman, Harris, & Orey, 2001). Similarly, numerous embedded validity indicators have been the subject of significant research and related reviews (e.g., Babikian, Boone, Lu, & Arnold, 2006; Heinly, Greve, Bianchini, Love, & Brennan, 2005; Iverson & Tulsky, 2003).

Methods of evaluating *self-report*. Individuals with and without brain disorders may complain of diminished abilities. For example, it is common for depressed individuals to complain of decreased concentration and memory abilities (e.g., Otto et al., 1994). Conversely, patients with severe TBI may minimize their deficits due to limited awareness. No examiner in any discipline is required to simply accept self-reported facts and history of examinees. The validity of *self-reported* disability and symptoms needs to be evaluated, especially when such complaints occur in a forensic context. When evaluating the validity of self-report, if possible, clinicians include measures that possess an internal means of assessing response bias.

Disorder-specific inventories. Interest in specific conditions that can become chronic, such as pain and post-traumatic stress disorder (PTSD), has generated a small but growing number of narrowly designed inventories that are intended to compare an individual's responses to those of patients who have been given the same diagnosis (e.g., PTSD) or who experience the same type of symptom (e.g., pain). Some of these inventories solicit information pertaining to abilities. In order to be considered appropriate for use, it is recommended that these instruments have validity scales that produce acceptable classification statistics, including sensitivity, specificity, positive predictive power, and negative predictive power. In general, disorder-specific inventories and symptom checklists that do not contain effective means for determining response bias and possible response invalidity should not be used in isolation. If, during an examination, self-report measures containing effective validity scales show clear and consistent invalid responding, data from self-report instruments that have no validity scales should not be relied on.

General personality inventories. The fields of clinical psychology and clinical neuropsychology have a long history of using general personality inventories that have a strong research foundation and well-established validity scales. These inventories are not equivalent in application to all conditions, requiring that clinicians keep abreast of relevant research. In the presence of signs of obvious response bias, whether negative or positive, clinicians should be appropriately conservative in interpreting clinical scales from these inventories. In the presence of response bias that is strong enough to invalidate responses on clinical scales, the clinical results of the inventory should not be interpreted.

Effort and malingering as applied to assessment of abilities

There is consensus regarding the existence of malingering. There is also consensus that the term malingering is appropriate for use in some instances. There is consensus regarding the meaning of significantly below-chance findings and what has been referred to as a ''compelling inconsistency'' (Bianchini et al., 2005), and both are viewed as individually reflecting a deliberate attempt to misrepresent one's abilities for which there are no alternative explanations. Intent may also be inferred as a result of the combined improbability of events, rather than a single definitive indication of intent (Larrabee, Greiffenstein, Greve, & Bianchini, 2007). Evidence that may be considered in the process of a differential diagnosis involving the possibility of malingering of abilities includes gross (1) disparity between real-world observations and either test performance or self-report, (2) inconsistency between type or severity of injury and test performances, (3) inconsistency between an individual's behavior when he/she is aware of being evaluated versus when not aware of not being evaluated, and (4) inconsistency across serial testings that cannot be explained by an underlying neurological process or known psychiatric condition. Whether a clinician is comfortable using the term ''malingering'' or not, the decision as to whether a case reflects malingering should have a basis in scientific research and related peer-reviewed assessment approaches.

The importance of evaluating performance validity when assessing abilities

Because psychometric indicators of performance invalidity are typically set to reduce false positive outcomes, positive findings (i.e., when credible thresholds have been surpassed) are more meaningful than negative findings. That is, a positive finding tends to ''rule in'' insufficient effort, whereas a negative finding may or may not ''rule out'' insufficient effort. However, not all instances of insufficient effort are indicative of malingering (Slick et al., 1999; Sweet, 1999). For example, a single indication of insufficient effort within a large battery of tests could indicate a transient effort problem that might not lead to the general conclusion that malingering is present, which is a decision based more broadly on all available information.

Sufficient effort is needed on every ability test in order to produce valid results. Moreover, effort is not a static concept, but rather a dynamic phenomenon that may vary throughout the examination. As such, *ideally*, effort should be evaluated repeatedly, if not continuously, throughout the course of an examination (Boone, 2009). When inconsistent or variable effort is shown to be present at any point during an evaluation, a reasonable and conservative conclusion is that all performances and obtained test scores may underestimate actual abilities, even those that occurred during apparent periods of adequate effort.

When the term malingering is applied to ability performances or self-report of abilities, it is not expected that malingering automatically or necessarily explains all of the individual's behaviors and presentations that are being scrutinized. For example, even though an examinee is found not to be credible on ability

testing, findings of depression on personality inventories may still be accurate. However, these other data may need to be viewed with caution and corroborated by additional outside sources of information. Moreover, performances on various neuropsychological tests used in a particular examination in which non-credible performances are present are best viewed as a lower bound estimate of actual level of ability (cf. Bush et al., 2005). Last, when the term malingering is applied, its use does not presume to explain all previous or subsequent behavior. The determination of the degree to which malingering is wholly, or only partly, explanatory is specific to the current evaluation.

Clinicians should be well apprised of the substantial relevant literature and be able to judge when it is appropriate to conclude that evidence of malingering is present. Whether a conclusion of malingering is made or not, when there is evidence of negative response bias there is consensus that determinations such as *non-credible*, *invalid*, and *implausible* can be made reliably on the basis of test results when they are viewed in a clinical context. When present, such invalid performances preclude the use of those data as a basis for: (1) opinions with regard to attribution to the cause at issue (e.g., accident, injury), (2) the nature and extent of possible deficits and disability, and (3) guiding treatment or evaluating treatment effectiveness.

Whether the context of a neuropsychological evaluation is clinical or forensic, the validity of self-report of abilities is an important source of information. Self-report information solicited in formal questionnaires and obtained by interviewing and review of historical records needs to be considered in evaluating response bias and malingering of abilities. When available to evaluate the validity of self-report, psychometric instruments that provide quantification of response validity relevant to the claimed disabilities are preferred. For example, a normal score on a validity scale created for detection of *feigned mental illness* is not necessarily informative regarding the credibility of cognitive or physical abilities; a normal score on a validity scale for detection of *feigned cognitive complaints* may be informative. That is, malingering of psychological conditions can occur independently of malingering of abilities, such as cognitive abilities (e.g., Nelson, Sweet, Berry, Bryant, & Granacher, 2007), which necessitates using appropriate assessment methods that can evaluate both, if psychological disorder *and* decreased ability claims are present.

Unlike psychometric indicators of invalid performance, objective standards for the evaluation of inconsistencies between the clinical presentation and evidence of capacities outside the clinical setting may not exist. In the absence of such standards, a determination that an inconsistency indicates invalidity should be cautious and conservative and in some cases may be left to the trier-of-fact. Inconsistencies in the self-report of individuals, including those related to (1) the severity of initial injury that increases across time, (2) inaccurate reporting of premorbid capacities or premorbid health, and (3) the evolution of clinical presentation across time when a static condition would be expected would raise suspicion of, but alone may not indicate, malingering.

In making related decisions, clinicians consider the known degree of self-report reliability in routine clinical settings, which is at times lacking. When feasible, it can be important to also consider collateral sources of information, to the degree

that the persons providing such information have access to the salient facts needed and are disinterested parties.

Documentation within reports of procedures used to assess abilities

Neuropsychologists routinely list in their reports the assessment methods and procedures, including standardized tests, that are utilized in their evaluations. Symptom validity measures and related procedures should also be listed in neuropsychological reports. Including the names of the symptom validity measures and embedded indicators on which opinions are based assists readers in understanding the bases of the conclusions and facilitates re-evaluations. Neuropsychologists use their own judgment when determining the degree of detail that is included in descriptions of validity measures, mindful of the importance of safeguarding specific information regarding these measures that would preclude valid use in the future, if specific information was to be disseminated to non-neuropsychologists.

Consideration of evaluation context

The base rate of negative response bias varies as a function of setting. Individuals presenting as litigants, defendants, or claimants in a criminal, civil, or disability proceeding or otherwise with motive to appear symptomatic (e.g., academic accommodations, drug seeking, excusing from military duty) show an increased risk (e.g., Greve et al., 2009; Mittenberg et al., 2002). For this reason, individuals seen in a forensic context should be given measures that will assist in identifying or ruling out response bias (cf. Bush et al., 2005), which is an expectation that has been supported for years by numerous forensic researchers and experts. Especially because research has shown repeatedly that experienced experts are inaccurate in identifying valid versus invalid ability performances from mere observation of behavior or test scores (e.g., Ekman, O'Sullivan, & Frank, 1999; Faust, 1995; Faust, Hart, Guilmette, & Arkes, 1988), for a clinician to choose not to use effort tests and embedded validity indicators requires a solid justification, especially within a forensic context. In fact, there is consensus that a *decision* not to use effort tests and embedded validity indicators would only rarely be justified (cf. Bush et al., 2005). It is more common to not be *able* to use such measures, such as when the evaluation is severely restricted in terms of time constraints or administrative prohibition (e.g., Social Security disability evaluation) or the individual being evaluated is not appropriate to be given such measures (e.g., severe and well-documented mental retardation precludes formal testing procedures). When multiple validity indicators are not or cannot be relied on, it is the clinician's responsibility to document the reasons and explicitly note any resulting limitations to interpretations of findings.

Response bias may occur in routine clinical and medical referrals, when no forensic context is evident. When clinicians are evaluating a *patient* who by virtue of claimed injuries is reasonably likely to become a *litigant* or *claimant*, the clinician

should consider the increased risk of insufficient effort and response bias and construct evaluations accordingly.

Attempts should be made to limit false positive identifications of response bias. More diagnostic certainty is required in contexts that involve higher relative costs for false positive errors (e.g., cases that potentially involve the death penalty). Toward this end, the clinician should be knowledgeable regarding the populations on which specific effort indicators were developed. To the extent that an examinee differs from test-normative and other comparison samples, the clinician should be appropriately cautious in drawing conclusions regarding the presence of response bias. When cultural, ethnic, and/or language factors are present that are known to affect the results of the instrument being used, clinicians need to adjust their thresholds of identifying insufficient effort and response bias accordingly (Salazar, Lu, Wen, & Boone, 2007). Although stand-alone effort tests have not generally been found to be impacted by age and education, at extremes of these variables, effort test failures have been known to occur (e.g., <third grade reading level on Green's Word Memory Test; Green & Flaro, 2003).

Consensus recommendations for practitioners related to assessment of abilities

The following points are viewed as important for neuropsychologists to consider when addressing response bias and malingering related to abilities:

- Use of psychometric indicators is the most valid approach to identifying neuropsychological response validity.
- Stand-alone effort measures and embedded validity indicators should both be employed.
- In their reports, neuropsychologists list the symptom validity measures and procedures that are utilized in evaluations. Clinicians explain the bases of their opinions to the extent required by the forensic context, while avoiding inclusion of specific information pertaining to these measures that could preclude valid future use.
- The evaluation of self-reported symptoms is best accomplished using psychometric instruments containing proven validity measures.
- Substantial inconsistencies between test data and "real-world" activities and between self-report and historical records should be considered. When integrating various sources of information, clinicians should be mindful of incomplete or false history, which when substantially present may reflect negative response bias.
- As risk relates to the setting in which the evaluation is taking place, clinicians should be mindful of the larger context of the evaluation and the potential for litigation to develop.
- As with all types of psychological assessment, neuropsychologists routinely are expected to encourage optimal effort as a means of attaining best performance.

- Substantial discrepancy between test results and those known to occur with the alleged medical or psychiatric disorder should raise concern regarding the presence of insufficient effort, response bias, and malingering.
- Because effort can vary during an evaluation, if possible clinicians should use multiple validity measures covering multiple domains distributed throughout the testing. If the circumstances are such that testing must be brief (e.g., Social Security disability evaluations), minimally, embedded effort indicators should be examined. When multiple validity indicators cannot be relied on, it is the clinician's responsibility to document the reasons and explicitly note the interpretive implications.
- As the number and extent of findings consistent with the absence or presence of response bias increases, confidence in conclusions regarding the validity of the examination is strengthened accordingly.
- Clinicians should be cognizant regarding when examinee characteristics do not match those of effort test-normative and comparison samples, and should adjust interpretations and choose measures accordingly.
- When a psychological disorder (e.g., depression) *and* ability deficits (e.g., memory) are claimed, clinicians should administer measures that can evaluate response bias related to both.
- Serial evaluations can be particularly helpful in discriminating between genuine injury and unrealistic performances or variable self-report of deficits and disabilities that reflect variable effort and/or response bias.

Consensus recommendations for future scientific investigation related to assessment of abilities

As with all other scientifically based practice specialties, there are ongoing research efforts that strive to improve the effectiveness of procedures used by neuropsychologists. The following topics are viewed as important and desirable to pursue in terms of gathering additional scientific knowledge relative to the assessment of response bias and malingering of abilities:

- Additional research is needed regarding the manner in which "weighting" or "aggregation" of cognitive effort measures and embedded validity indicators of ability measures is best accomplished in clinical practice.
- Populations at risk of failing effort and embedded validity indicators despite best effort should be investigated.
- Additional methods should be developed that will provide evidence of *deliberate intent* to feign.
- New ability tests should have validity indicators created at the time of test construction.
- Cost–benefit of response bias assessment and malingering detection methods should be analyzed.
- Effort measures and embedded validity indicators should be applied to pediatric samples.
- Identification of the point at which risk of response bias has effectively been ruled out should be investigated.

SOMATIC ISSUES

Conceptual and operational definitions

Response validity regarding somatic symptoms relates to excessive subjective disability attributed to somatic dysfunction, which may result from claims of injury or illness. Examples include claims of physical symptoms purported to be associated with fibromyalgia, chronic fatigue syndrome, multiple chemical sensitivity, and other medically unexplained conditions (Binder & Campbell, 2004), in which exaggeration and/or fabrication of symptoms and disability is suspected to play a role.

Somatic complaints may be either specific or nonspecific in nature. Specific complaints may include complaints of focal weakness or paresthesia, lateralized sensorimotor impairment, and primary sensory changes (e.g., loss or reduction of vision, smell, or other sensory abilities). Non-specific complaints may include general malaise, fatigue, generalized weakness, dizziness, balance problems or vague generalized complaints of ''tingling'' or numbness. A primary domain of somatic complaints is pain (e.g., headache, low back pain, neck pain), which may be specific or non-specific. Pain symptoms may vary in terms of the qualitative nature of the pain, as well as pain severity, and the subjective report of pain-related disability (Bianchini et al., 2005).

Non-credible somatic symptoms may present as exaggerated or atypical symptom report on general personality tests, or exaggerated or atypical symptom report on specialized rating scales. Persons complaining of non-credible somatic symptoms may under-perform on measures of strength/dexterity, and may also show evidence of impaired cognitive performance that is atypical for bona fide clinical disorders. Evidence of non-credible somatic disability presentation shows base rates of 30–40% in secondary gain contexts (Greve, Ord, Bianchini, & Curtis, 2009; Mittenberg et al., 2002; Meyers, Millis, & Volkert, 2002) similar to those reported for non-credible cognitive performance disability.

Methods of assessment

Self-report. As noted in the Ability section of this consensus statement, it is important to assess the reliability and validity of examinee's self-report that is obtained by interviewing. Comparison of an examinee's self-reported history with available information from other reliable sources can assist in determining whether or not history is accurate and symptoms are consistently reported. Clinicians should be mindful of the known tendency among some forensic examinees to misrepresent (usually in a positive direction) their pre-incident historical status, in terms of cognitive, somatic, and/or psychological function (e.g., Gunstad & Suhr, 2001; Mittenberg, DiGuilio, Perrin, & Bass, 1992).

Extended personality inventories—e.g., Minnesota Multiphasic Personality Inventory-2 (MMPI-2); Personality Assessment Inventory (PAI)—as well as other more focused pain scales—e.g., Pain Patient Profile (P3)—are relevant to the assessment of potentially over-reported somatic symptoms. Such inventories typically show the greatest exaggeration on scales related to somatic symptoms,

with additional exaggeration, to a smaller degree, on scales measuring symptoms of depression and anxiety (Larrabee, 1998, 2007). Specialized validity scales, such as the MMPI-2 FBS (Symptom Validity scale) of the MMPI-2, have shown good discriminative validity and classification accuracy in the detection of exaggerated somatic symptoms (Greiffenstein, Fox, & Lees-Haley, 2007; Nelson, Sweet, & Demakis, 2006). Newer scales, (e.g., Fs; Ben-Porath & Tellegen, 2008), specifically tailored to somatic over-reporting have also been developed and may show merit in detecting exaggerated symptom presentation.

Motor skills. Unlike self-report, examination of motor skill performance in non-credible somatic symptom report is less studied. However, select investigations suggest potential utility of motor skill assessment in the detection of feigned somatic presentations. Finger-tapping scores among compensation-seeking individuals with independent evidence of non-credible cognitive performance fall substantially below the clinical benchmarks relative to non-compensation-seeking *patients* with genuine neurological injury (Arnold et al., 2005; Larrabee, 2003). Abnormally poor performance on a grip strength device can also provide evidence of feigned motor impairment (Greiffenstein, 2007). Last, feigned motor impairment can occur in disease-deficit-incompatible patterns. For instance, poor performance on gross relative to fine motor tasks is a pattern opposite to that produced by patients with documented neurologically based motor dysfunction (e.g., Greiffenstein, Baker, & Gola, 1996). Clinical benchmarks of motor performance among patients with known motor dysfunction (e.g., Butters, Goldstein, Allen, & Shemansky, 1998; Greiffenstein, 2007) should be considered.

Sensory/perceptual. A number of authors (Binder, Kindermann, Heaton, & Salinsky, 1998; Binder, Salinksy, & Smith, 1994; Mittenberg, Rotholc, Russell, & Heilbronner, 1996; Trueblood & Schmidt, 1993) have summarized difficulties in evaluating finger localization and graphesthesia in the context of non-credible impairment of sensoriperceptual function. In select cases that involve feigned primary sensory impairment, customized forced-choice measures can inform assessment of non-credible presentation (Pankratz, 1979).

Symptom validity tests. Patients exaggerating somatic symptoms may also fail standard cognitive symptom validity tests, also referred to as effort tests. In fact research has shown that base rates of symptom validity test failure are increased among compensation-seeking individuals with conditions, such as fibromyalgia (Gervais et al., 2001), toxic exposure (van Hout, Schmand, Wekking, Hageman, & Deelman, 2003), and chronic pain (Meyers & Diep, 2000; Meyers & Volbrecht, 2003). Legitimate somatic discomfort does not lead to patterns of suspected feigned cognitive impairment (Etherton, Bianchini, Greve, & Ciota, 2005). For these reasons, inclusion of cognitive symptom validity tests (forced-choice or non-forced-choice approaches) should be strongly considered. See the Ability Issues section of this statement for additional discussion of symptom validity tests.

Salient variables

Demand characteristics of a given response validity measure vary according to the nature of the external incentive. For instance, consistent with the scale's original

development in 1991, FBS appears more relevant to civil settings (Nelson et al., 2006) because over-reporting of severe psychiatric symptoms is more characteristic of individuals evaluated in criminal settings (Wygant et al., 2007).

Response validity on many ability measures does not relate to gender. However, a subset (e.g., response validity of motor performance; Arnold et al., 2005) has been found to benefit from gender-based cut scores. Similarly, assessing the validity of symptom reporting for some methods may benefit from consideration of gender, whereas other methods may not benefit from analysis of this or other moderator variables. There are limited data with regard to the effect of moderator variables on other validity scales and symptom validity tests designed to evaluate the accuracy of somatic symptom complaints.

Consensus recommendations for practitioners related to assessment of somatic symptoms

The following points are viewed as important for neuropsychologists to consider when addressing response bias and malingering of somatic presentation:

- When assessing for non-credible somatic presentation, use multiple well-validated measures covering domains of self-report, performance, and symptom validity. As the number and extent of findings consistent with absence or presence of response bias increase, confidence in conclusions regarding the validity of the examination is strengthened accordingly.
- Carefully rule out plausible alternative explanations, other than malingering, for the somatic presentation, as it is critically important to keep false positives to a minimum. The clinician is encouraged to consider actuarial data along with clinical judgment of patient self-report when making determinations of veracity of somatic complaints.
- Related to assessing the validity of somatic complaints, history information should be evaluated for completeness and accuracy.
- Keep current with literature that addresses non-credible somatic presentation.

Consensus recommendations for future scientific investigation related to Assessment of somatic symptoms

The following points are viewed as important to consider when addressing response bias and malingering of somatic presentation:

- It is essential that researchers seriously consider the *criterion problem*. That is, researchers must form criterion groups of "malingerers" independently of dependent variables.
- It is recommended that researchers employ stringent inclusion/exclusion criteria in their investigations. Control or comparison groups should consist of medical patients without known external incentives. The ideal criterion group is one that has (a) little to no clinical or laboratory evidence for pathology, combined with (b) symptom histories that are compellingly illogical to a reasonable clinician.

- Many past investigations have employed, as a comparison group, patients who themselves are seeking or receiving compensation, without any screening for symptom exaggeration or invalid test performance. Should an investigator wish to examine medical patients in whom external incentive is a factor, comparison group participants should be carefully screened a priori for the presence of non-credible somatic symptoms.
- In conjunction with other more commonly employed approaches, future studies might consider use of alternative functional capacity measures (e.g., Waddell signs; Waddell, McCulloch, Kummel, & Venner, 1980) as converging supportive evidence of exaggerated somatic presentation. Colleagues from other disciplines, such as medicine, might collaborate in establishing optimal classification accuracy results in such areas as functional capacity assessment, although neuropsychological researchers will need to appreciate limitations and differences in these approaches as compared to more typical exaggeration and comparison groups.
- The influence of demographic variables, such as gender, should continue to be a focus of investigation.

PSYCHOLOGICAL ISSUES

Conceptual and operational definitions

"Psychological issues" refers to those psychological/psychiatric disorders or conditions that may be seen in evaluations of claimants in secondary gain contexts. Typical contexts may include Independent Medical Examinations (IMEs) referred by insurance carriers, personal injury cases in civil litigation, disability evaluations for workers' compensation and Social Security, Department of Veterans Affairs claims for disability compensation, criminal prosecution for competency to proceed, the insanity defense/diminished capacity, or mitigation of the death penalty, among others (Boyd, McLearen, Meyer, & Denney, 2007; Denney, 2007, 2008; Greiffenstein, 2007; Rogers, 2008).

In this section we discuss response validity assessment in claims of psychopathology and/or emotional distress, specifically in the form of exaggeration, symptom promotion, or frank symptom fabrication. Common disorders primarily include, but are not limited to, anxiety disorders, depression and related disorders (e.g., bipolar disorder), psychoses, and post-traumatic stress disorder (PTSD). The committee recognizes that PTSD is classified as a type of anxiety disorder (APA, 2000), but chose to represent it as a separate diagnostic entity within this consensus statement because of its prevalence in disability claims, especially within military settings and the Department of Veterans Affairs, and also because of the presence of unique attributes of the disorder (i.e., the etiology is dependent on a traumatic stressor; APA, 2000). The current list is not intended to be exhaustive. Other common disorders may also be encountered in secondary gain examination contexts, such as attention deficit hyperactivity disorder (ADHD) and/or learning disability (LD) within an assessment context for test-taking accommodations (Mapou, 2008; Osmon & Mano, 2009). Some disability insurance carriers may also compensate individuals with substance abuse disorders.

Diagnostic considerations

Regardless of the specific type of psychopathology, it is incumbent upon the examiner to pursue a formal, in-depth diagnostic evaluation. It is the *totality* of the claimant's presentation that should be taken into account when assessing the validity of claims of psychopathology and/or emotional distress. Neuropsychologists are well suited to this endeavor by virtue of their scientific training and experience in psychopathology, knowledge of differential diagnosis, expertise in psychometrics, and skills in the assessment of symptom validity and response bias. In fact, neuropsychologists have largely been responsible for the development and growth of this important area (Boone, 2007; Larrabee et al., 2007; Sweet, 2009; Sweet, Ecklund-Johnson, & Malina, 2008). Examiners must familiarize themselves with the discrete diagnostic criteria of the condition in question with the most current authoritative diagnostic sources. Specific attention should be directed toward the most common features of the disorder in question, including the onset, course, symptom picture, co-morbidities, treatment efforts, and response to treatment. With these diagnostic criteria in mind, the examiner can address whether the claimant's presentation is, or is not, compatible with what is known about the disorder. The presence of inconsistencies and contradictory data should be noted and documented.

Diagnostic considerations necessarily take into account the following information.

Onset. The compatibility of disorder onset with the general principles of what is known about the particular psychological/psychiatric condition in question is important. In this regard, the examiner may note obvious inconsistencies, such as the claim of "schizophrenia" with onset well past the typical age of onset. In criminal contexts, in individuals with no psychiatric history, observations of first symptom onset *after the defendant's arrest* should raise suspicion. The examiner must determine whether the onset of the disorder makes sense and is consistent with well-established knowledge of psychopathology.

Similarly, elucidation of the *claimant's and his/her family history* can be informative. Again, as stated in earlier sections of this consensus statement, self-report information obtained by interviewing needs to be evaluated for accuracy. Some early childhood circumstances (e.g., abuse, neglect) have the potential to predispose an individual to the development of psychopathology later in life (Grover et al., 2007; Molnar, Buka, & Kessler, 2001). Although the lack of a family history of psychopathology may not undermine the credibility of claimant's presentation, the presence of a positive family psychiatric history may ultimately be helpful in differential diagnoses.

Symptom presentation. The consistency of reported and observed symptoms with what is typically expected in the disorder can be informative. The presence of atypical/unusual symptoms, such as odd hallucinations or those that appear to represent a layman's conception of the disorder, may raise suspicion of symptom fabrication. The presence of an unusual combination of symptoms (e.g., a claimant who experiences hallucinations only in the presence of his/her pet dog) may suggest symptom fabrication (Guy, Kwartner, & Miller, 2006).

Course. The degree to which symptom presentation follows a course or pattern over time that is typical or atypical of the disorder at issue can be revealing. For example, a claimant purporting to have auditory hallucinations 24 hours a day (i.e., without even brief periods of remission) or the claimant who is catatonically depressed at a given time only to be spontaneous and conversant the next moment may raise questions regarding validity of symptoms.

Treatment/response to treatment. An examinee's documented history may indicate that the claimant has been formally diagnosed with the disorder in question by an appropriately credentialed mental health professional. This documented history within the examinee's records conceivably may provide some historical evidence of credibility. However, the accuracy and ultimate credibility of that diagnosis should be viewed with an appropriate degree of skepticism and confirmed within the broad and comprehensive scope of the present examination, rather than necessarily be taken at face value. Thus, the possibility of incorrect prior diagnoses should be carefully explored. It is important to note whether the diagnosis was made in a clinical context apart from a secondary gain context. Has the examinee sought appropriate treatment for the disorder? Has he/she responded as expected to treatment efforts or does he/she remain largely refractory? In this regard, there is little doubt of the meaningfulness of the situation when an examinee remains refractory to treatment for depression only *after* application for compensation.

Assessment methodology

General considerations. The assessment of psychological and psychiatric issues by neuropsychologists is a data-driven enterprise, which may best be characterized as a search for "consistencies and inconsistencies." Operationalizing this concept, the examiner strives to answer the question, "Does what I am learning about this claimant make sense in light of the putative claim, the diagnosis, history and totality of the presentation?" Assessment is a multi-faceted endeavor requiring the integration of numerous discrete data sources, typically including: (1) review of records; (2) psychosocial history obtained by interviewing; (3) observations of the claimant's behavior during the assessment period; (4) consideration of information from collateral sources, such as significant others, employers, etc., when available and appropriate; (5) formal psychological/neuropsychological testing; (6) response validity assessment procedures; and (7) surveillance video/audio, when available. Examiners are encouraged to utilize best practices and follow current practice guidelines (e.g., AACN, 2007).

When feasible, it is often appropriate to obtain all available medical, psychiatric, legal, and other relevant records. Careful scrutiny of records, with particular attention to premorbid status, reported onset and course of symptoms, treatment efforts, and response to treatment can be informative. This review may reveal the presence of contradictory and/or confirmatory opinions of other professionals, which warrant consideration.

The examiner's interview of the claimant will necessarily be thorough, with particular attention paid to possible inconsistencies in behavior, demeanor,

symptoms, etc. Is the claimant's presentation consistent with the reported diagnosis and aspects of the history? Does his or her mental status make sense in light of reported symptoms and history, or does it stand at odds with all or part of it? Is the onset of the disorder plausible? Some examinees, particularly in criminal forensic evaluative settings, may be so naïve and unsophisticated as to "develop" symptoms after their arrest (Morgan, 2008). While most claimants are likely to be more sophisticated than that, the examiner should be sensitive to such temporal factors.

Historical information can be relevant to all forms of psychopathology, particularly with regard to post-traumatic stress disorder (PTSD), as the nature and extent of the trauma allegedly experienced by the claimant is crucial in determining a plausible etiology and diagnosis of the disorder. Examiners attempt to inquire about all aspects of the alleged stressful event(s) in question, the context of occurrence, and with special inquiry into its plausibility. Some claimants may present a history in which they did not experience an event that "involved actual or threatened death or serious injury, or threat to the physical integrity of self or others" (APA, 2000, p. 467). Claimants may sometimes present plausible-sounding events, which may ultimately be shown to be fabricated or significantly embellished. A detailed history of pre- and post-injury life stressors should be explored with emotional stress claims.

Cultural, ethnic, and socio-economic factors should be taken into assessment and diagnostic considerations, as some forms and/or expressions of psychopathology may be varied or more prevalent among certain demographic groups (U.S. Department of Health & Human Services, 2000). Distinct clinical presentations and/or prevalence of psychopathology may differ among some cultures and ethnic groups (Draguns & Tanaka-Matsumi, 2003).

Consistency/inconsistency of examinee behavior over time can be important. This is within an evaluation and also true as it relates to an examinee's behavior outside of the test environment, which might be revealed via video surveillance. When present, the presence of unusual symptoms, an irregular pattern or course of the disorder, vague or odd onset, non-credible stressors, atypical response to treatment efforts, a waxing-waning/remitting course, etc, need to be considered. Collateral information can reveal consistency or inconsistency in history and presentation and should be obtained and reviewed when possible. Such information may confirm or fail to confirm the claimant's self-report.

Psychological and neuropsychological assessment

Because many claims regarding psychiatric diagnoses involve claimant reports of impaired cognitive functions (ostensibly related to the psychiatric condition), comprehensive neuropsychological assessment is frequently warranted. Most claims related to psychological/psychiatric disorders include alleged impairments in attention, processing speed, memory, and executive functions. Examinees may report cognitive dysfunction in the area of attention deficits and/or memory impairment reportedly related to anxiety, depression, or some other psychiatric disorder. In addition, some symptoms that primarily present as a neurocognitive disorder are believed to have their basis in psychopathology (e.g., psychogenic non-epileptic seizures, PNES; Alper, 1994; Reuber & Elger, 2003), thus seemingly

clouding the sometimes artificial distinction between cognitive and psychiatric disorders and symptoms. As with any cognitive assessment, particularly in a secondary gain context, formal tests of cognitive and emotional symptom validity are indicated. Examiners are advised not to rely on a single symptom validity test, but multiple measures administered throughout the assessment day(s) are suggested. Such measures may be useful to assist in the determination of non-credible claims of emotional distress, exaggerated claims of cognitive dysfunction and/or poor effort. See the Ability Issues and Somatic Issues section of this statement for additional discussion on symptom validity tests.

In the assessment of exaggerated symptoms of emotional/psychiatric distress and the detection of feigned psychological symptoms, a number of freestanding measures are available for various conditions (e.g., PTSD, depression). Multi-scale inventories of psychopathology (e.g., MMPI-2) are well known to neuropsychologists. Many of these include specific scales and items designed to assess the validity of complaints of psychiatric and emotional symptoms. An example of this type of measure is the Structured Interview of Reported Symptoms (SIRS).

Clinicians need to stay current with the relevant peer-reviewed literature in order to determine whether to rely on such scales. Use of multiple indices of both cognitive and emotional validity in assessments can insure that the diagnostic process is as accurate as possible.

Neuropsychologists, by virtue of their training, clinical judgment, and experience, and informed by well-validated assessment methods, knowledge of the condition in question, and relevant scientific literature, can identify the presence and severity of genuine psychopathology. The presence of symptom exaggeration, response bias, and/or symptom fabrication can be determined by the presence of a combination of factors, including non-credible/invalid response patterns, non-credible history/presentation, and or lack of corroboration by collateral data sources and/or medical records.

A determination of malingered psychopathology, whether exaggerated or wholly feigned, can be made on the basis of self-report, psychological/ neuropsychological test results, and observations/ corroborative information. When claims of emotional distress are disproportionate to the history provided by the claimant and other sources of information, and/or when inconsistencies in self-report over time and across evaluations are present, and/or when symptom reports do not follow what is known about the progression of symptoms and the natural history of the claimed disorder, and/or when symptom presentation is atypical of the disorder, a determination of exaggerated or feigned psychopathology may be justified. It is important to recognize that even in cases of well-confirmed, genuine psychopathology, malingering may co-exist. Indeed, genuine psychopathology and symptom fabrication are not mutually exclusive (Morgan & Gervais, 2009; Morgan, Millis, & Mesnik, 2009; Slick et al., 1999). However, when neuropsychological and psychological test results reveal multiple indicators of invalid test performance, even in the context of genuine psychopathology, there is an increased possibility that exaggerated and or feigned/malingered psychopathology is present.

Consensus recommendations for practitioners related to assessment of psychological symptoms

The following points are viewed as important for neuropsychologists to consider when addressing response bias and malingering of psychological symptoms of psychological presentation:

- Clinicians engaged in assessment of an individual's psychological issues within a forensic context need to be aware that self-report may be biased, false, or incomplete, and proactively evaluate this possibility.
- For individuals undergoing forensic assessment, both cognitive and/or emotional complaints are common co-occurrences in this population, requiring that both domains be appropriately assessed and the validity of each determined.
- Clinicians would benefit from establishing their own practice-specific database on examinees, consisting of in-depth interview and complete history, observations of behavior, informant/collateral interviews, and formal neuropsychological/psychological testing.
- Assessment instruments and scales should be utilized that provide the most current, scientifically informed methodology for the assessment of emotional/psychopathology and cognitive factors.
- Clinicians should utilize multiple symptom validity measures administered throughout the evaluation. Confidence in conclusions is expected to be commensurate with the number and extent of findings demonstrating absence or presence of response bias.
- Because the co-occurrence of genuine psychopathology and feigned/exaggerated symptoms is common, examiners should, as much as possible, attempt to delineate the relative presence of each. Examiners should be familiar with base rates of mental disorders and symptoms in the general population.
- Clinicians are encouraged to use best clinical practices, assessment methodologies (instruments, scales, scoring criteria, etc.) that are current, and be current consumers of the relevant scientific literature.
- Because research and clinical experience indicate that some cultural and ethnic differences exist with regard to the presentation of psychopathology, clinicians are encouraged to consider such factors, as appropriate to the individual case.

Consensus recommendations for future scientific investigation related to assessment of psychological symptoms

The following points are viewed as important for neuropsychologists to consider when addressing response bias and malingering:

- Researchers are encouraged to continue to investigate emotional and cognitive response validity issues in known psychiatric groups and to continue the development of new instruments and scales to detect response bias.
- The relationship between genuine psychopathology and invalid response patterns requires ongoing empirical investigation and scientists are encouraged to work toward better elucidation of this relationship.

- Scientists engaged in personality/psychopathology research are encouraged to become more familiar with invalid self-report and attempt to understand those individual differences that may contribute to distortion, exaggeration, or feigning of self-report.

RESEARCH EVIDENCE AND SCIENTIFIC ISSUES

Research designs

Three research designs that have been most commonly used in effort-testing studies are: (a) simulation; (b) criterion groups (''known'' groups); and (c) differential prevalence designs. The reader is referred to Rogers (2008) for a detailed description and analysis of these different designs.

In analog simulation studies, non-clinical participants, such as community volunteers or college students, are assigned to different experimental conditions (e.g., given instructions to simulate) and compared to clinical samples of participants with verified cognitive impairment on the test of interest. This analog design often provides tight experimental control as well as a practical and cost-effective method for examining ''proof of concept'' for new tests. A major limitation of this design is one of generalizability (i.e., that the behavior elicited may be different than behavior encountered in the real world; Rogers, 2008).

A criterion (''known groups) design uses a priori criteria to define response bias, such as selecting people in litigation for mild brain injury claims who perform at chance level on a forced-choice test. Criteria typically are chosen that optimize a very low false positive rate. As with the analog simulation design, the criterion group's performance on the test of interest is compared to a clinical group's performance. The criterion group design has the strength of clinical relevance by including people who have real-world incentives to malinger. However, developing appropriate external criteria for defining response bias can be a major methodological challenge.

The differential prevalence design assumes that response bias prevalence varies as a function of environmental context and likelihood of incentives to malinger. For example, it is inferred that response bias is more common in persons who are litigating versus those who are not. Groups are composed on the basis of assumed incentives. Use of differential prevalence designs may be most useful in initial validation studies. A major limitation of the differential prevalence design is that it is difficult, if not impossible, to know the accuracy of the inferences made.

It is the consensus of this panel that the simulation and criterion groups designs represent rigorous and clinically relevant research designs. The differential prevalence design can yield additional information relevant to test development, but it should not be used as the sole or primary research design for test validation. In general, researchers are encouraged to use multiple designs during test development and validation. In addition, it is important to include appropriate clinical groups in a systematic program of research that is designed to examine a test's construct and predictive validity.

Statistical and methodological issues

Single versus multiple indicators. Of the various types of tests designed to detect response bias, single, *stand alone* forced-choice tests have been the best validated. These single tests will continue to be useful in neuropsychological assessment. However, sole reliance on single tests raises diagnostic concerns. The extent to which single tests are vulnerable to detection, disclosure, and coaching will determine whether a particular test will lose diagnostic sensitivity to response bias (Horwitz & McCaffrey, 2006; Youngjohn, 1995). In addition, there is the question of whether the detection of response bias can be improved incrementally beyond the use of single tests. Researchers have encouraged the use of multiple measures as a method for increasing classification accuracy (Larrabee, 2003; Nelson et al., 2003; Victor, Boone, Serpa, Buehler, & Ziegler, 2008).

The panel recommends more research on methods that combine multiple stand-alone tests and/or embedded measures. Common methods include unit weighting (Dawes, 1979) and logistic regression (Menard, 2002). Multivariable test composites may be more resistant to coaching, especially when based on standard neuropsychological tests. In addition, methods like logistic regression can assist in determining which tests or variables produce incremental power in differentiating groups, and estimate optimal weights for the tests, while controlling for test redundancy. Simple unit weighting has been criticized for failing to account for the dependencies among tests, which likely results in overly optimistic predictions (Chan, Deeks, Macaskill, & Irwig, 2008). However, in their review of the literature, Bobko, Roth, and Buster (2007) found that unit weighting has substantial predictive validity. Larrabee (2008) demonstrated the usefulness of a unit-weighting method in the symptom validity context.

Although multivariable composites offer considerable potential in deriving new tests and embedded measures, there are certain statistical pitfalls that need to be avoided. When developing multivariable models, investigators must select a set of predictor variables. It is not uncommon to have dozens of variables from which to choose. However, one risks overfitting the model if there are not at least 10 to 20 participants for every predictor variable in the regression context. In the case of binary logistic regression, this estimate is based on the sample size of the smaller of the two groups, not the total sample (Harrell, 2001). Overfitting will exaggerate the predictive worth of the model. The chance of finding spurious associations between the dependent variable and predictor variables will be greatly increased. Some investigators have looked to stepwise variable selection to solve the problem of too many variables. Stepwise variable selection includes a variety of methods for selecting subsets of variables on the basis of the change in the residual sum of squares as a result of including or excluding the variable from the model. However, Harrell (2001) has noted that: (a) stepwise techniques produce biased standard errors and inflated R^2 values, (b) the F and χ^2 test statistics reported by stepwise methods do not have the claimed distributions, (c) the reported p values are too small; (d) multicollinearity among the predictor variables greatly affects which variables are selected by stepwise methods; and (e) the final model will often contain variables that are not accurate predictors of the dependent variable.

Harrell (2001) provides detailed strategies for variable selection. First, research hypotheses, theory, and past research findings should still guide variable selection. Newly developed statistical techniques such as Bayesian model averaging (Wang, Zhang, & Bakhai, 2004) and penalized maximum likelihood estimation (Moons, Donders, Steyerberg, & Harrell, 2004) represent different, but useful approaches, to the variable selection problem.

Whether using a multivariable composite or a single test, neuropsychologists should not rely on single, fixed cut scores. Neuropsychologists appreciate and consider a range of cut scores and associated diagnostic test statistics in choosing the cut score to be applied to a specific case. The decision-making process occurs in different contexts, such that the relative costs of false positive and false negative errors will not be constant across situations. Raising or lowering a test's cut score will increase or decrease the test's sensitivity and specificity in an inverse fashion: when sensitivity is increased, specificity decreases. To assist clinicians in this decision-making process, investigators, journal editors, and test publishers are strongly encouraged to provide a broad range of cut scores with their respective diagnostic test statistics (e.g., sensitivities, specificities, and likelihood ratios).

All tests and score combinations require validation. There are two major methods of model validation: external and internal (Harrell, 2001). External validation generally involves an independent validation sample with participants from a different, but similar, population (Justice, Covinsky, & Berlin, 1999). Although ideal, external validation can be costly in terms of time and resources. Some investigators will split their original sample into derivation and validation samples. However, this procedure is problematic because it can result in lower precision and power (Harrell, 2001). Internal validation procedures, such as bootstrapping (Steyerberg, Bleeker, Moll, Grobbee, & Moons, 2003; Steyerberg et al., 2001), offer a useful supplement to external validation and do not require sample splitting. The basic concept of bootstrapping involves drawing samples (e.g., 500) with replacement from the original data set of the same size, computing the statistic of interest, and examining how the statistic changes over the 500 repetitions. This process allows the investigators to determine the degree of "optimism" in the original model.

Diagnostic statistics. Differences on test scores between two groups are commonly analyzed statistically with the *t*-test or the Mann-Whitney-Wilcoxon test. Although these tests are useful, they are insufficient to evaluate the performance of tests and embedded measures. Researchers and clinicians need to have a working knowledge of common diagnostic statistics, including but not necessarily limited to, those listed below (Straus, Richardson, Glasziou, & Haynes, 2005).

- Sensitivity: The proportion of persons with the disorder (or condition or behavior of interest) who obtain a positive score on the effort test (i.e., fail the test).
- Specificity: The proportion of persons without the disorder (or without the condition or behavior of interest) who obtain a negative score on the effort test (i.e., pass the effort test).
- Positive predictive value (PPV): The proportion of patients with positive test results who are correctly classified. It should be noted that PPV, unlike

sensitivity and specifity, varies as a function of the prevalence (i.e., base rate) of the disorder (or condition or behavior of interest).

$$PPV = \frac{(sensitivity)(prevalence)}{(sensitivity)(prevalence) + (1 - specificity)(1 - prevalence)}$$

- Negative predicted value (NPV): The proportion of patients with negative test results who are correctly classified. NPV also varies as a function of disorder prevalence.

$$NPV = \frac{(\text{specificity})\,(\text{prevalence})}{(\text{specificity})(prevalence) + (1 - \text{sensitivity})(1 - \text{prevalence})}$$

- Likelihood ratio: [Sensitivity / (1 – Specificity)]. The percentage of people with the disorder with a positive test (i.e., true positives) divided by the percentage of people without the disorder who have a positive test results (i.e., false positives). It indicates how many times more (or less) likely it is that people with the disorder obtain a positive test result compared to those without the disorder who obtain a positive test result. A likelihood ratio greater than 1.0 indicates that the test result is associated with the presence of the disorder of interest. A likelihood ratio of less than 1.0 is associated with the absence of the disorder (Straus et al., 2005). Likelihood ratios from 2.0 to 5.0 yield small increases in the post-test probability, from 5.0 to 10.0 moderate increases, and above 10.0 large increases (Grimes & Schulz, 2005).
- Area under the curve (AUC): Area under the receiver operating characteristic (ROC) curve, or AUC. The ROC curve is a plot of a test's sensitivity (plotted on the y-axis) versus its false positive rate (i.e., 1 – specificity) plotted on its *x*-axis. Each point on the graph is generated by a different cut score on the test. Each point can be connected, a curve can be generated, and the area under the curve can be estimated parametrically (via maximum likelihood estimation) or non-parametrically (via the trapezoidal rule). The ROC curve area has several interpretations (Zhou, Obuchowski, & McClish, 2002): (a) the average value of sensitivity for all possible values of specificity, (b) the average value of specificity for all possible values of sensitivity, and (c) the probability that a randomly selected patient with the disorder has a test result indicating greater suspicion than that of a randomly selected patient without the disorder. Hosmer and Lemeshow (2000) offer some guidelines for interpreting the magnitude of AUC: (a) .50 = no discrimination; (b) .70 – .80 = acceptable discrimination; (c) .80 – .90 = excellent discrimination; and (d) ≥.90 = outstanding discrimination.

Reporting of diagnostic accuracy studies

It is recommended that investigators who are reporting the findings of diagnostic studies of SVTs follow the Standards for Reporting of Diagnostic

Accuracy (STARD) Guidelines (see Bossuyt et al., 2003). In 1999, the Cochrane Diagnostic and Screening Test Methods Working Group discussed the poor quality and substandard reporting of many diagnostic test evaluations. This working group then developed the STARD initiative. The STARD Guidelines are designed to improve the quality of the reporting of diagnostic studies. A flow diagram and spreadsheet are available to assist investigators in describing the important components of study design, method of participant recruitment, description of the reference and index tests, the conduct of the study, the administration of the tests, and the results. The diagram and checklist are available on the STARD website: http://www.stard-statement.org/. Neuropsychologists are encouraged to apply the STARD guidelines to evaluate diagnostic studies of embedded measures, stand-alone tests, and multivariable algorithms.

SUMMARY

Issues related to effort, response bias, and malingering have been prominent in the professional journals and at the professional meetings of clinical neuropsychologists for years. The detection of problematic effort and response bias, and the determination that malingering is present or absent, are important functions of clinicians, particularly because there is a substantial risk that such factors could be present among the numerous cases assessed by neuropsychologists in a secondary gain context. Even in a routine clinical context the presence of problematic effort and response bias can potentially invalidate results. The assessment of effort and genuine reporting of symptoms is important in all evaluations. The present Consensus Conference Statement is intended to assist clinicians and researchers, and should be viewed as reflecting current knowledge and available instruments. However, this area of investigation is by no means static. With the degree of energy and interest invested in these topics by clinicians and researchers, a dynamic and continued evolution of concepts, definitions, and related assessment procedures is to be expected. Practitioners are encouraged to remain informed regarding the local rules governing provision of expert opinions regarding response bias, effort, and malingering.

REFERENCES

Alper, K. (1994). Nonepileptic seizures. *Neurology Clinics*, *12*, 153–173.

American Academy of Clinical Neuropsychology (AACN) (2007). American Academy of Clinical Neuropsychology (AACN) Practice Guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist*, *21*, 209–231.

American Psychiatric Association (APA) (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision). Washington, DC: APA.

Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S., et al. (2005). Sensitivity and specificity of finger-tapping test scores for the detection of suspect effort. *The Clinical Neuropsychologist*, *19*, 105–120.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, *20*, 145–159.

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

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, *5*, 404–417.

Binder, L. M., & Campbell, K. A. (2004). Medically unexplained symptoms and neuropsychological assessment. *Journal of Clinical and Experimental Neuropsychology*, *26*, 369–392.

Binder, L. M., Kindermann, S. S., Heaton, R. K., & Salinsky, M. C. (1998). Neuropsychological impairment in patients with nonepileptic seizures. *Archives of Clinical Neuropsychology*, *13*, 513–522.

Binder, L. M., Salinksy, M. C., & Smith, S. P. (1994). Psychological correlates of psychogenic seizures. *Journal of Clinical and Experimental Neuropsychology*, *16*, 524–530.

Bobko, P., Roth, P., & Buster, M. (2007). The usefulness of unit weights in creating composite scores. *Organizational Research Methods*, *10*, 289–709.

Boone, K. B. (Ed.) (2007a). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.

Boone, K. B. (2007b). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment*. New York: Guilford Press.

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist*, *23*, 729–741.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., et al. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clinical Radiology*, *58*, 575–580.

Boyd, A. R., McLearen, A. M., Meyer, R. G., & Denney, R. L. (2007). *Detection of deception*. Sarasota, FL: Professional Resource Press.

Bush, S. S., Ruff, R. M., Troster, A., Barth, J., Koffler, S. P., Pliskin, N. H., et al. (2005). NAN position paper: Symptom validity assessment: Practice issues and medical necessity. *Archives of Clinical Neuropsychology*, *20*, 419–426.

Butters, M. A., Goldstein, G., Allen, D. N., & Shemansky, W. J. (1998). Neuropsychological similarities and differences among Huntington's disease, multiple sclerosis, and cortical dementia. *Archives of Clinical Neuropsychology*, *13*, 721–730.

Chan, S. F., Deeks, J. J., Macaskill, P., & Irwig, L. (2008). Three methods to construct predictive models using logistic regression and likelihood ratios to facilitate adjustment for pretest probability give similar results. *Journal of Clinical Epidemiology*, *61*, 52–63.

Committee on Ethical Guidelines for Forensic Psychologists (1991). Specialty guidelines for forensic psychologists. *Law and Human Behavior*, *15*, 655–665.

Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.

Delis, D. C., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology*, *22*, 589–604.

Denney, R. L. (2007). Assessment of malingering in criminal forensic neuropsychological settings. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 428–452). New York: Guilford Press.

Denney, R. L. (2008). Negative response bias and malingering during neuropsychological assessment in criminal forensic settings. In R. L. Denney & J. P. Sullivan (Eds.), *Clinical neuropsychology in the criminal forensic setting* (pp. 91–134). New York: Guilford Press.

Draguns, J. G., & Tanaka-Matsumi, J. (2003). Assessment of psychopathology across and within cultures: Issues and findings. *Behavior Research and Therapy*, *41*, 755–776.

Ekman, P., O'Sullivan, M., & Frank, M. (1999). A few can catch a liar. *Psychological Science*, *10*, 263–266.

Etherton, J., Bianchini, K., Greve, K., & Ciota, M. (2005). Test of Memory Malingering is unaffected by laboratory-induced pain: Implications for clinical use. *Archives of Clinical Neuropsychology*, *20*, 375–384.

Faust, D. (1995). The detection of deception. *Neurologic Clinics*, *13*, 255–265.

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

Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, *14*, 3–11.

Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology*, *28*, 1892–1899.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology*, *9*(3), 189–207.

Greiffenstein, M. F. (2007). Motor, sensory, and perceptual-motor pseudoabnormalities. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 100–130). New York: Oxford University Press.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. *Journal of the International Neuropsychological Society*, *2*, 477–485.

Greiffenstein, M. F., Fox, D., & Lees-Haley, P. R. (2007). The MMPI-2 Fake Bad Scale in detection of non-credible brain-injury claims. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment. A neuropsychological perspective*. New York: Guilford Press.

Greve, K. W., Binder, L. M., & Bianchini, K. J. (2009). Rates of below-chance performance in forced-choice symptom validity tests. *The Clinical Neuropsychologist*, *23*, 534–544.

Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). The prevalence of malingering in chronic pain patients referred for psychological evaluation in a medico-legal context. *Archives of Physical Medicine & Rehabilitation*, *90*, 1117–1126.

Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet*, *365*, 1500–1505.

Grover, K. E., Carpenter, L. L., Price, L. H., Gagne, G. G., Mello., A. F., Mello., M. F., et al. (2007). The relationship between childhood abuse and adult personality disorder symptoms. *Journal of Personality Disorders*, *21*, 442–447.

Gunstad, J., & Suhr, J. A. (2001). 'Expectation as Etiology' versus 'The Good Old Days': Postconcussion syndrome symptom reporting in athletes, headache sufferers, and depressed individuals. *Journal of the International Neuropsychological Society*, *7*, 323–333.

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.

Harrell, F. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag.

Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. L., & Brennan, A. (2005). WAIS Digit Span-based indicators of malingered neurocognitive dysfunction: Classification accuracy in traumatic brain injury. *Assessment*, *12*, 429–444.

Hom, J., & Denney, R. (Eds.) (2002). *Detection of response bias in forensic neuropsychology*. New York: Haworth Medical Press.

Horwitz, J. E., & McCaffrey, R. J. (2006). A review of Internet sites regarding independent medical examinations: Implications for clinical neuropsychological practitioners. *Applied Neuropsychology*, *13*, 175–179.

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley Interscience.

Iverson, G., & Tulsky, D. (2003). Detecting malingering on the WAIS-III: Unusual Digit Span performance patterns in the normal populations and in clinical groups. *Archives of Clinical Neuropsychology*, *18*, 1–9.

Justice, A. C., Covinsky, K. E., & Berlin, J. A. (1999). Assessing the generalizability of prognostic information. *Annals of Internal Medicine*, *130*, 515–524.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist*, *12*, 179–188.

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

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

Larrabee, G. J. (2007b). Evaluation of exaggerated health and injury symptomatology. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 264–286). New York: Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*, 666–679.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–371). New York: Oxford University Press.

Mapou, R. L. (2008). Learning disabilities in adults. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (pp. 696–728). New York: Psychology Press.

McCrea, M., Pliskin, N., Barth, J., Cox, D., Fink, J., French, L., et al. (2008). Official position of the military TBI task force on the role of neuropsychology and rehabilitation psychology in the evaluation, management, and research of military veterans with traumatic brain injury. *The Clinical Neuropsychologist*, *22*, 10–26.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Meyers, J. E., & Diep, A. (2000). Assessment of malingering in chronic pain patients using neuropsychological tests. *Applied Neuropsychology*, *7*, 133–139.

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

Meyers, J. E., & Volbrecht, M. E. (2003). Validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, *18*, 261–276.

Mittenberg, W., DiGiulio, D. V., Perrin, S., & Bass, A. E. (1992). Symptoms following mild head injury: Expectation as aetiology. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*, 200–204.

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*, 1094–1102.

Mittenberg, W., Rotholc, A., Russell, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead-Reitan Battery. *Archives of Clinical Neuropsychology*, *11*, 271–281.

Molnar, B. E., Buka, S. L., & Kessler, R. C. (2001). Child sexual abuse and subsequent psychopathology: Results from the National Comorbidity Survey. *American Journal of Public Health*, *91*, 753–760.

Moons, K. G., Donders, A. R., Steyerberg, E. W., & Harrell, F. E. (2004). Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: A clinical example. *Journal of Clinical Epidemiology*, *57*, 1262–1270.

Morgan, J. E. (2008). Non-credible competence: How to handle "newbies," "wannabees," and forensic "experts" who know better or should know better. In R. L. Heilbronner (Ed.), *Neuropsychology in the courtroom: Expert analysis of reports and testimony* (pp. 53–65). New York: The Guilford Press.

Morgan, J. E., & Gervais, R. O. (2009). Definite malingering or probable malingering: Multidimensional symptom exaggeration in a case of depression. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 122–131). New York: Psychology Press.

Morgan, J. E., Millis, S. R., & Mesnik, J. (2009). Malingered dementia and feigned psychosis. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 231–243). New York: Psychology Press.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *Clinical Neuropsychologist*, *17*, 263–272.

Nelson, N., Sweet, J., Berry, D., Bryant, F., & Granacher, R. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society*, *13*, 440–449.

Nelson, N. W., Sweet, J. J., & Demakis, G. J. (2006). Meta-analysis of the MMPI-2 Fake Bad Scale: Utility in forensic practice. *The Clinical Neuropsychologist*, *20*, 39–58.

Osmon, D. C., & Mano, Q. (2009). Malingered attention deficit hyperactivity disorder: Effort, depression and dependence in the pursuit of academic accommodations. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 386–398). New York: Psychology Press.

Otto, M. W., Bruder, G. E., Fava, M., Delis, D. C., Quitkin, F. M., & Rosenbaum, J. F. (1994). Norms for depressed patients for the California Verbal Learning Test: Associations with depression severity and self-report of cognitive difficulties. *Archives of Clinical Neuropsychology*, *9*, 81–88.

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*, *43*, 421–422.

Reuber, M., & Elger, C. E. (2003). Psychogenic nonepileptic seizures: Review and update. *Epilepsy & Behavior*, *4*, 205–216.

Rogers, R. (Ed.) (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York: Guilford Press.

Rogers, R. (2008). Researching response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed.). New York: The Guilford Press.

Romero, H. R., Lageman, S. K., Kamath, V., Irani, F., Sim, A., Suarez, P., et al. (2009). Proceedings: Challenges in the Neuropsychological Assessment of Ethnic Minorities: A Problem Solving Summit. *The Clinical Neuropsychologist*, *23*, 761–779.

Salazar, X., Lu, P., Wen, J., & Boone, K. (2007). The use of effort tests in ethnic minorities and non-English-speaking and English as a second language populations. In K. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 405–427). New York: Guilford Press.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.

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

Steyerberg, E. W., Bleeker, S. E., Moll, H. A., Grobbee, D. E., & Moons, K. G. (2003). Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*, *56*, 441–447.

Steyerberg, E. W., Harrell Jr, F. E., Borsboom, G. J., Eijkemans, M. J., Vergouwe, Y., & Habbema, J. D. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, *54*, 774–781.

Straus, S., Richardson, W., Glasziou, P., & Haynes, R. (2005). *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). New York: Elsevier Churchill Livingstone.

Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice* (pp. 255–285). New York: Psychology Press.

Sweet, J. J. (2009). Appendix: Forensic bibliography: Effort/malingering and other common forensic topics encountered by clinical neuropsychologists. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 566–630). New York: Psychology Press.

Sweet, J. J., Ecklund-Johnson, E., & Malina, A. (2008). Forensic neuropsychology: An overview of issues and directions. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (pp. 869–890). New York: Psychology Press.

Sweet, J. J., King, J. H., Malina, A. C., Bergman, M. A., & Simmons, A. (2002). Documenting the prominence of forensic neuropsychology at national meetings and in relevant professional journals from 1990 to 2000. *The Clinical Neuropsychologist*, *16*, 481–494.

Thompson, G. B. (2002). The Victoria Symptom Validity Test: An enhanced test of symptom validity. *Journal of Forensic Neuropsychology*, *2*, 43–67.

Tombaugh, T. N. (2002). The Test of Memory Malingering (TOMM) in forensic psychology. *Journal of Forensic Neuropsychology*, *2*, 69–96.

Trueblood, W., & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *15*, 578–590.

U.S. Department of Health, Human Services (HHS) (2000). *Mental health: A report of the surgeon general*. Rockville, MD: HHS.

van Hout, M. S., Schmand, B., Wekking, E. M., Hageman, G., & Deelman, B. G. (2003). Suboptimal performance on neuropsychological tests in patients with suspected chronic toxic encephalopathy. *Neurotoxicology*, *24*, 547–551.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*, 45–73.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2008). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*, 297–313.

Waddell, G., McCulloch, J. A., Kummel, E., & Venner, R. M. (1980). Nonorganic physical signs in low-back pain. *Spine*, *5*, 193–203.

Wang, D., Zhang, W., & Bakhai, A. (2004). Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. *Statistics in Medicine*, *23*, 3451–3457.

Wygant, D. B., Sellbom, M., Ben-Porath, Y. S., Stafford, K. P., Freeman, D. B., & Heilbronner, R. L. (2007). The relation between symptom validity testing and MMPI-2 scores as a function of forensic evaluation context. *Archives of Clinical Neuropsychology*, *22*, 489–499.

Youngjohn, J. (1995). Confirmed attorney coaching prior to neuropsychological assessment. *Assessment*, *2*, 279–283.

Zhou, X., Obuchowski, N., & McClish, D. (2002). *Statistical methods in diagnostic medicine*. New York: Wiley Interscience.

APPENDIX: LIST OF PRE-CONFERENCE READINGS

General (for all groups to use)

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic considerations for malingered neuro-cognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.

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

Sweet, J. J. (2009). Appendix: Forensic bibliography: Effort/malingering and other common forensic topics encountered by clinical neuropsychologists. In J. Morgan & J. Sweet (Eds.), *Neuropsychology of malingering casebook*. New York: Psychology Press.

Heilbronner Working Group: Definitions and Differential Diagnosis

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* [sections on malingering and somatoform disorders]. Washington, DC: APA.

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment*. New York: Guilford Press.

Greve, K., & Bianchini, K. (2004). Setting empirical cut-offs on psychometric indicators of negative response bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychology*, *19*, 533–541.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

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

Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice*. New York: Psychology Press.

Sweet Working Group: Ability Issues

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P, Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy and planning committee. *Archives of Clinical Neuropsychology*, *20*, 419–426.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist*, *17*, 263–272.

Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society*, *13*(3), 440–449.

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*, 213–224.

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

Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, *16*, 495–505.

Larrabee Working Group: Somatic Issues

Bianchini, K. B., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *Spine Journal*, *5*, 404–417.

Lanyon, R. I. (2003). Assessing the misrepresentation of health problems. *Journal of Personality Assessment*, *81*, 1–10.

Larrabee, G. J. (1998). Somatic malingering on the MMPI and MMPI-2 in personal injury litigants. *The Clinical Neuropsychologist*, *12*, 179–188.

Larrabee, G. J. (2003). Exaggerated pain report in litigants with malingered neurocognitive dysfunction. *The Clinical Neuropsychologist*, *17*, 395–401.

Nelson, N. W., Sweet, J. J., & Demakis, G. J. (2006). Meta-analysis of the MMPI-2 Fake Bad Scale: Utility in forensic practice. *The Clinical Neuropsychologist*, *20*, 39–58.

Rohling, M. L., Binder, L. M., & Langhinrischen-Rohling, J. (1995). Money matters: A meta-analytic review of the association between financial compensation and the experience and treatment of chronic pain. *Health Psychology*, *14*, 537–547.

Morgan Working Group: Psychological Issues

Delis, D. C., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology*, *22*, 589–604.

Demakis, G. J., Gervais, R. O., & Rohling, M. L. (2007). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of PTSD. *The Clinical Neuropsychologist*, *21*, 1–17.

Iverson, G. L., Le Page, J., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist*, *21*, 532–546.

Greiffenstein, M. F., et al. (2004). The Fake Bad Scale and MMPI-2 family in detection of implausible psychological trauma claims. *The Clinical Neuropsychologist*, *18*, 573–590.

Greiffenstein, M. F., & Baker, J. W. (2007). Validity testing in dually diagnosed post-traumatic stress and mild closed head injury. *The Clinical Neuropsychologist*, *21*, 1–18.

Rohling, M. L., Green, P., Allen, L. M., & Iverson, G. L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology*, *17*, 205–222.

Millis Working Group: Research Evidence and Scientific Issues

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., et al. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Clinical Radiology*, *58*(8), 575–580.

Larrabee, G., Greiffenstein, M. F., Greve, K., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Millis, S. R. (2003). Statistical practices: The seven deadly sins. *Child Neuropsychology*, *9*(3), 221–233.

Rutjes, A. W., Reitsma, J. B., Coomarasamy, A., Khan, K. S., & Bossuyt, P. M. (2007). Evaluation of diagnostic tests when there is no gold standard. *A review of methods*. Health Technology Assessment, *11*(50), iii, ix–51.

Straus, S., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2005). Diagnosis and screening. In S. Straus, W. S. Richardson, P. Glasziou, & R. B. Haynes (Eds.), *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). London: Elsevier.

Wilkinson, L., & Task Force on Statistical Inference APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604.