Psychol. Inj. and Law
DOI 10.1007/s12207-015-9218-2

Ethical Quandaries for Psychologists in Workers' Compensation Settings: the GAF Gaffe

Shadi Gholizadeh · Vanessa L. Malcarne · Michael E. Schatman

Received: 13 January 2015 / Accepted: 27 January 2015

Abstract Psychologists working within the forensic realm of workers' compensation (WC) evaluative settings can be confronted with a host of unique, ethical quandaries worthy of discussion. The ethical challenges presented by the use of one specific assessment instrument, the Global Assessment of Function Scale (GAF), a clinician-rated, single, numeric scale used as a global assessment of an individual's psychological, social, and occupational functioning, in WC settings are explored. Reliability and validity of the GAF are discussed in order to evaluate whether its use as a single indicator of psychiatric permanent disability for WC determinations is psychometrically, and subsequently ethically, justified. The present analysis demonstrates that psychologists working in evaluative contexts in WC settings may be putting themselves in ethically precarious situations in their legally mandated use of the GAF to evaluate permanent disability relating to alleged psychiatric injuries. The dearth of psychometric support to justify the use of the GAF to determine psychiatric impairment suggests that the current practice is ethically, psychometrically, and clinically problematic. The authors provide recommendations for more robust assessment procedures.

Keywords Workers' compensation · Permanent disability · Psychiatric injury · Disability assessment · Global Assessment of Functioning

S. Gholizadeh (✉) · V. L. Malcarne
Joint Doctoral Program in Clinical Psychology, SDSU/UC San Diego,
6363 Alvarado Court, Suite 103, San Diego, CA 92120-4913, USA
e-mail: shadigholizadeh@gmail.com

V. L. Malcarne
Department of Psychology, San Diego State University, San Diego, CA, USA

M. E. Schatman
Foundation for Ethics in Pain Care, Bellevue, WA, USA

Published online: 13 February 2015

Background

While psychologists and psychiatrists working in any mental health-care setting are often faced with ethically challenging situations, those working within the forensic realm of workers' compensation (WC) evaluations can be confronted with a host of unique ethical quandaries worthy of discussion. In an effort to contextualize the complexity of modern WC systems and underscore the necessity for a system that compensates injured workers, Guyton (1999) traces a history of WC dating back to approximately 2050 B.C. via the law of Ur in ancient Sumeria that described an ancient compensation schedule mapped onto specific worker injuries. Homologues existed throughout the ancient world that were often quite detailed, for example the valuing of a thumb joint as one half the value of a finger in ancient Arab edicts (Guyton, 1999). The primary concern of modern systems has shifted from protecting injured workers from destitution to also taking into consideration the protection of employers from frivolous lawsuits (Schatman, 2012). While initially confined to industrial accidents, the domain of WC has broadened to include less acute physical injuries, such as overuse injuries and, more recently, psychiatric injuries. However, efforts to more clearly define the ever-growing roles of mental health professionals working in WC systems have not matched the burgeoning presence of psychologists in these spheres.

There are myriad potential situations in which the goals and ethical duties of a psychologist may conflict with those of the parties involved in a given WC case. For example, questions around what party should be considered the psychologist's "client" (e.g., the injured worker; the insurance carrier), concerns about releasing test materials and/or data to parties requesting this information, and uncertainties about the bounds of confidentiality and how these should be addressed in the consenting process are several among the many oft-encountered ethical challenges. However, the authors of the present paper will focus on the ethical challenges presented by the use of one specific assessment

instrument, the Global Assessment of Functioning (GAF) Scale, a clinician-rated, single, numeric scale used as a global assessment of an individual's psychological, social, and occupational functioning, for psychologists who are serving evaluative roles. Challenges will be considered within the context of the American Psychological Association (APA) Ethical Principles of Psychologists and Code of Conduct (subsequently referred to as the APA Ethics Code; APA, 2010). As it is beyond the scope of the present article to provide a description of each nation's and state/province's laws and guidelines, California will be used as a case study throughout this paper, with a special emphasis on the state's reliance on the GAF for the evaluation of psychiatric impairment in WC settings. California has served as a case example in other papers (e.g., Schwartz, 1993), because of the state's relatively high WC expenditures and history of controversy and "politicking surrounding workers' compensation in California" (p. 985). However, the issues raised will be of relevance to all psychologists working in forensic settings in which assessment guidelines are vague, controversial, or problematic.

Psychologists in Workers' Compensation Settings: a California Case Study

The Changing Face of California Workers' Compensation

Prior to delving into the specific professional and ethical issues associated with the use of the GAF, a brief history of WC in California is warranted. WC in California was formally established in 1913 (CHSWC, 2008). A number of important reforms have been initiated since that time, with the most recent being the enactment of Senate Bill 863 (SB 863) in 2013, a far-reaching WC reform package (discussed below) that included substantial limitations to permanent disability (PD) compensation for psychiatric injuries. By virtue of being a state that offers permanent partial disability (PPD) benefits such that greater severity of injury confers entitlement to higher benefits, California needed a systematic means of ranking the severity of alleged impairments. This ranking is referred to as the permanent disability ranking (PDR) in California (Reville, Seabury, Neuhauser, Burton, & Greenberg 2005). Prior to the 2004 passage of Senate Bill 899, the California WC system was considered one of the most controversial in the USA, largely because of its reliance on a system for assessing disability that was under fire for breeding inconsistency and inviting fraud (Reville et al., 2005). Senate Bill 899 provided a means to systemize the determination of PDR, via the American Medical Association Guides, 5th Edition (hereafter, AMA Guides), for *physical* disabilities. Following an evaluation of impairment using the AMA Guides, the impairment rating is transformed into a disability rating from which benefits are determined following adjustment for age, occupation, and future earning capacity (FEC) diminishment. SB 863 further standardized PDR via abolishment of the tailored FEC adjustment such that all injuries after January 1, 2013, are adjusted by a standard factor of 1.4. Thus, the trend has been toward moving away from the controversial waters of subjective disability rating protocols onto the *terra firma* of objective rating schemes. However, the zeitgeist of objectivity afforded to physical disability in WC has not been extended to psychiatric disability.

Prior to 2005, the Methods of Measurement of Psychiatric Disability [see CAL. CODE REGS. tit. 8, § 43 (2009), http://www.dir.ca.gov/t8/43.html.] enacted in 1992 dictated that psychiatric impairment and subsequent disability be evaluated per eight work functions and their manifestations (see Table 1). Clinicians were first asked to list all disabling symptoms that an individual reported and then required to rate the level to which symptoms led to impairment of the

Table 1 Work function impairment form for evaluation of mental and emotional impairment used for injuries prior to 2005

Work function	Example functional manifestation
1. Ability to comprehend and follow instruction	• The ability to maintain attention and concentration for necessary periods
2. Ability to perform simple and repetitive task	• The ability to perform activities of a routine nature
3. Ability to maintain a work pace appropriate to a given work load	• The ability to perform activities within a schedule, maintain regular attendance, and be punctual
4. Ability to perform complex and varied tasks	• The ability to perform a variety of duties, often changing from one task to another of a different nature without loss of efficiency or composure
5. Ability to relate to other people beyond giving and receiving instructions	• The ability to get along with co-workers or peers
6. Ability to influence people	• The ability to interact appropriately with others
7. Ability to make generalizations, evaluations, or decisions without immediate supervision	• The ability to recognize potential hazards and follow appropriate precautions
8. Ability to accept and carry out responsibility for direction, control, and planning	• The ability to set realistic goals or make plans independently of others

Adapted directly from the impairment form provided via the DIR. Note that the actual form provides more examples of functional manifestations than provided in this adapted table

aforementioned eight work functions on a 5-point response scale ranging from minimal ("discomfort, but not disabling") to severe ("unable to perform work function"). An updated schedule for rating permanent disabilities effective January 1, 2005 changed the way that psychologists rate psychiatric impairment for PDR purposes [see LABOR AND WORKFORCE DEVELOPMENT AGENCY, DEPARTMENT OF INDUSTRIAL RELATIONS, DIVISION OF WORKERS' COMPENSATION, SCHEDULE FOR RATING PERMANENT DISABILITIES UNDER THE PROVISIONS OF THE LABOR CODE OF THE STATE OF CALIFORNIA (2005)]. California mandated use of the GAF in determining psychiatric impairment and assigning disability in California's WC reform of 2004. Specifically, psychologists are asked to assign a numerical GAF rating, which is then converted to a whole person impairment (WPI) rating using a conversion system provided in the Schedule for Permanent Disabilities. Adjustments to the WPI are made for (1) occupation type, (2) age, and, prior to SB 863, (3) FEC. Of note, the FEC adjustments were empirically based on data from a comprehensive RAND Corporation analysis that justified the highest adjustments made for psychiatric disabilities (California Division of Workers' Compensation, 2005). The law states, "…a psychiatric impairment receives a higher FEC adjustment because RAND data shows (sic) that a relatively high wage loss corresponds to the average psychiatric standard permanent disability rating" (p. 1–6). However, as a consequence of SB 863, the differential FEC adjustment was abolished for psychiatric disabilities as well.

Thus, starting on January 1, 2005, the GAF virtually usurped the aforementioned, lengthier evaluative process covering the different work functions and their manifestations such that the single numerical GAF rating became a deciding factor in determining the value of a WC settlement or award. Consequently, once "predominant causality" of an alleged psychiatric injury has been shown to arise from actual events of employment [see CAL. LAB. CODE § 3208.3 (West 2011)], the GAF is essentially one part of the ultimate determination of PD compensability, along with apportionment to work-related duties, as will be described further below.

Definitions and Clarifications

While psychologists can both diagnose and treat injured workers in the WC setting, the present focus is on psychologists working in evaluative contexts, specifically serving as medical evaluators. Qualified medical evaluators (QMEs) are defined by the Department of Industrial Relations (DIR) as "…qualified physicians who are certified by the Division of Workers' Compensation-Medical Unit to examine injured workers to evaluate disability and write medical-legal reports. The reports are used to determine an injured worker's eligibility for WC benefits. QMEs include….psychologists." [see LABOR AND WORKFORCE DEVELOPMENT AGENCY, DEPARTMENT OF INDUSTRIAL RELATIONS, DIVISION OF WORKERS' COMPENSATION, DWC QUALIFIED MEDICAL EVALUATOR (QME) PROCESS (2014)]. Although there are distinctions among different types of evaluative experts (e.g., agreed medical examiners, independent medical examiners) depending on the various selection processes used to identify the expert, for the purposes of the present analysis, the term QME will be used here.

Psychologists evaluating alleged psychiatric injury in California must (1) evaluate the individual to assess the presence of a "mental disorder which causes disability or need for medical treatment…diagnosed using the terminology and criteria of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Third Edition-Revised, or the terminology and diagnostic criteria of other psychiatric diagnostic manuals generally approved and accepted nationally by practitioners in the field of psychiatric medicine" [note that this language has not been updated to reflect more current revisions of the Diagnostic and Statistical Manual of Mental Disorders (DSM)]and (2) "demonstrate by a preponderance of the evidence that actual events of employment were predominant as to all causes combined of the psychiatric injury" [i.e., apportionment; see CAL. LAB. CODE § 3208.3 (West 2011)]. Predominance is generally understood as greater than 50 % in such settings [see *Dep't of Corr. v. Workers' Comp. Appeals Bd.*, 90 Cal.Rptr.2d 716, 720 (Ct. App. 1999)]; although injuries resulting from a violent act must only show "substantial" cause, which is generally understood as at least 35 to 40 % [see CAL. LAB. CODE § 3208.3(b)(3) (West, 2011); see also *Sonoma State Univ. v. Workers' Comp. Appeals Bd.*, 48 Cal.Rptr.3d 330, 332–34 (Ct. App. 2006)]. Once these two criteria have been satisfied, the determination of PD compensation rests solely on the GAF score assigned by the evaluator.

The requirements to become a QME as a psychologist are largely the same as those for other clinicians, with the exception of a specific requirement that psychologists demonstrate one of the following: (1) board certification in clinical psychology by the American Board of Professional Psychology and a minimum of 5 years of doctoral experience; (2) a doctoral degree in psychology recognized by the "Administrative Director" and at minimum 5 years of doctoral experience in the diagnosis and treatment of mental disorders; or (3) a minimum of 5 years of doctoral experience in the diagnosis and treatment of mental disorders and evidence of having served as an AME on eight or more occasions prior to January 1, 1990 [see Division of Workers' Compensation-Medical Unit: Application for Appointment as Qualified Medical Evaluator (10, 2013), https://www.dir.ca.gov/dwc/FORMS/QMEForms/QMEForm100.pdf]. In addition to passing a QME competency exam, a 12-hour course on disability evaluation report writing is required. It should be noted that nowhere in the requirements is explicit training in assessment generally or the GAF specifically mandated.

It is important to clarify that although "impairment" and "disability" are often used interchangeably, the role of the QME is to evaluate an alleged impairment such that disability can be determined. Per the AMA Guides, which serves as the key reference for California WC and is used to evaluate and rate impairment for physical injuries, impairment is "an alteration of an individual's health status; a deviation from normal…" (Cocchiarella & Anderson, 2001). Disability is defined by the AMA Guides as "an alteration of an individual's capacity to meet personal, social, or occupational demands because of an impairment" (Cocchiarella & Anderson, 2001). However, the relationship between disability and impairment is a complex one, and there are multiple competing definitions for each construct (Schultz, 2008). For example, in an elegant "disentangling" (p. 103) of the construct of disability in the realm of psychological injury, Schultz and Stewart (2008) explore six models of disability relevant to medicolegal contexts (e.g., biomedical, psychosocial) and the parallel return to work paradigms associated with each, further evincing the complexity of the issue. It is assumed that an injured worker's impairment will improve until a point of maximal medical improvement (MMI), which the California DIR also terms "permanent and stationary." The period from date of injury to MMI is termed the *temporary disability period*, whereas the period following MMI is the *permanent disability period* (Reville et al., 2005). Once MMI status has been reached, the clinician provides a GAF rating from which a Whole Person Impairment (WPI) score is directly calculated and on which PD compensation is based. GAF scores of 70 and higher are associated with 0 WPI.

Finally, there are various types of mental injuries to which a worker can be subject. The three types of mental injury in WC settings are the following: (1) physical-mental injuries, wherein "physical injuries lead to disabling psychological repercussions"; (2) mental-physical injuries, wherein "mental stimuli…result(s) in physical disabilities"; and (3) mental-mental injuries, wherein "mental stimuli…result(s) in a debilitating mental response" (Riley, 2000). Table 2 summarizes each type of claim and provides the relevant case reference. Although the "mental stress" claim is perhaps the most controversial, and most often perpetuated in the media as frivolous, it should be noted that, "traditionally, tort law and workers' compensation statutes denied recovery for claims of emotional distress unaccompanied by physical injury" (Matsumoto, 1994; p. 1328).

California has historically been among the more amenable states to these mental distress-type claims (i.e., psychological distress without accompanying physical injury). However, compensability of mental injuries occurring on or after January 1, 2013 in California has been significantly limited per the enactment of SB 863, and the updated status of each type of claim is also provided in Table 2. The primary changes of relevance to psychologists resulting from SB 863 come in the form of PD compensation for psychiatric disorders. Specifically, barring certain specific exceptions (see Table 2), workers claiming psychiatric disorders no longer receive PD compensation if it is deemed that the alleged psychiatric disorder is a consequence of a physical injury that is being compensated (i.e., physical-mental claim). As described in CAL. LAB. CODE § 4660.1(c)(1) (West 2011):

> …there shall be no increases in impairment ratings for sleep dysfunction, sexual dysfunction, or psychiatric disorder, or any combination thereof, arising out of a compensable physical injury. Nothing in this section shall limit the ability of an injured employee to obtain treatment for sleep dysfunction, sexual dysfunction or psychiatric disorder, if any, that are a consequence of an industrial injury.

Table 2 Summary of types of psychiatric injury claims

Claim type	Case reference[a]	Summary	Status following SB 863
Physical-mental	• Hohlstein v. St. Louis Roofing Co., 49 S.W.2d 226 (Mo. Ct. App. 1932; mental injury resulted from worker's 20-foot fall to the ground at a job site)	• Physical injury is precipitant to and deemed causal of mental injury	• Prohibits increase in impairment rating for sleep dysfunction, sexual dysfunction, or psychiatric disorders arising from physical conditions except if resulting from a "violent act" or "catastrophic injury" • Status of temporary treatment for these conditions unchanged
Mental-physical	• Montgomery County v. Grounds, 862 S.W.2d 35 (Tex. App. 1993; worker suffered heart attack after he was not informed that he would not be indicted for altering police reports)	• Mental injury is precipitant to and deemed causal of physical injury	• Unchanged
Mental-mental	• Bryant v. Giani Inv. Co., 626 So.2d 390 (La. Ct. App. 1993; worker suffered mental injury following verbal argument with supervisor)	• The injury *and* its effects are both mental in nature	Unchanged

[a] Case reference information adapted from Riley (2010)

Thus, to provide a highly simplified example for illustrative purposes, we can consider hypothetical cases of two workers who fall off a building in the course of employment duties, both of whom develop posttraumatic stress disorder (PTSD) following the fall. Employee A also suffers orthopedic injuries to her back while employee B does not. Employee B would be entitled to PD for the PTSD (a mental-mental industrial injury) while employee A, assuming that the orthopedic injuries are compensable, would likely not be entitled to PD for the PTSD, which could be argued as related to the physical injury (a physical-mental industrial injury).[1]

In summary, reforms in the California WC system have targeted limitation of compensability of psychiatric injuries over the years, although the aforementioned limitations imposed by SB 863 are the most limiting to date in terms of PD compensation for mental injury claims. SB 863 stipulates the additional evaluative task of demonstrating that an alleged mental injury for which an individual is seeking PD compensation is not related to a physical injury for which the individual is seeking compensation.

The limitations in compensability of psychiatric injuries imposed by SB 863are largely a response to complaints regarding the subjectivity of mental claims and the commonly held view that physical-mental claims are often frivolous "add-ons" or fodder for fraud or malingering. Such mental claims thus have developed a notoriety in the WC sphere as a tactic by unscrupulous applicant attorneys to leverage the threat of "expensive and protracted ligation" to strong-arm insurance companies to "'pay off" questionable stress claims" (Matsumoto, 1994, p. 1336). Unfortunately, rather than exploring means to better operationalize mental disability and reform assessment guidelines, the trend has simply been toward limitations on the compensability of mental claims.

The GAF: How a Single-Item Global Measure Became the Psychiatric Permanent Disability Rating Despot

A paramount contribution of the field of psychology has been the development of standardized, replicable, and psychometrically sound tools for diagnosing, evaluating, predicting, and otherwise measuring phenomena with which individuals present. From single-item, global symptom scales to complex multi-factor instruments designed to measure various aspects of personality, psychometric assessments provide psychologists with tools to understand the degree to which various constructs—e.g., depression, suicidality, quality of life, malingering—exist and evolve in individuals. The zeitgeist of the field of psychology emphasizes evidence-based treatments, of which evidence-based assessment is an integral component. Unfortunately, psychologists often fail to utilize psychometrically valid and reliable measures to substantiate their opinions in forensic settings (see Underwager & Wakefield, 1993, and Pope, Butcher, & Seelen, 2006, for examples).

As part of the previously described sweeping WC overhaul of 2004, California moved to the use of the AMA Guides for physical injuries and the GAF for psychiatric injuries, despite the fact that the AMA Guides specifically cautioned against using percentages for mental (i.e., psychiatric) impairments, stating that

> Percentages are not provided to estimate mental impairment…Unlike cases with some organ systems, there are no precise measures of impairment in mental disorders. The use of percentages implies a certainty that does not exist. Percentages are likely to be used inflexibly by adjudicators, who then are less likely to take into account the many factors that influence mental and behavioral impairment. In addition, the authors are unaware of data that show the reliability of the impairment percentages. (Cocchiarella & Anderson, 2001, p. 361)

The AMA Guides offered an alternative to a single numerical rating via a five-tiered rating (ranging from no impairment to extreme impairment) for the following four areas of function: activities of daily living, social functioning, concentration, and adaptation. However, the California WC system opted to utilize a numerical impairment rating scale by way of the GAF.

A Brief History of the GAF

The GAF is a clinician-rated assessment of an individual's psychological, social, and occupational functioning on a single-item response continuum ranging from 1 (i.e., representing "the hypothetically sickest individual") to 100 (i.e., "the hypothetically healthiest individual"; Startup, Jackson, & Bendix, 2002, p. 417). There is also an option to score 0 in instances in which a clinician believes that information is inadequate to assign an accurate score. The predecessor to the GAF was the Health-Sickness Rating Scale (HSRS), developed in 1962 by Luborsky as a global measure of mental health wherein clinicians judge an individual on seven mental health dimensions (e.g., patient's need to be protected, seriousness of symptoms, ability of the individual to utilize abilities in work and other settings, etc.). Clinicians using the HSRS were instructed in a series of instructions accompanying the measure to consider an individual's functioning across the seven dimensions in order to "find a region where a patient's condition might be located" on a 100-point scale ranging from 0 ("any condition which, if unattended, would

[1] However, Cal. Lab. Code § 4660.1(c)(2) does provide two caveats to the new restrictions in cases in which the compensable psychiatric injury was deemed to result from (1) the injured worker's status as victim of a violent act or (2) a catastrophic injury, such as the loss of a limb (see Table 2). It should be noted that both of these exceptions are extremely ambiguous and will likely remain contentious, inviting case law to more clearly operationalize terms such as "catastrophic."

quickly result in the patient's death, but not necessarily by his own hand") to 100 ("an ideal state of complete functioning integration, of resiliency in the face of stress, of happiness and social effectiveness"). To choose the specific numerical rating once a broad range was identified, clinicians were directed to a set of 34 previously ranked sample cases, to which they could compare the individual. Endicott, Spitzer, Fleiss, and Cohen (1976) revised the HSRS by dividing the 1–100-scale into ten, evenly distributed anchor points and eliminating the diagnostic examples that the HSRS provided along with the 34 specific case examples and replacing them with more clearly defined behavioral statements and general examples (e.g., "moderate symptoms…few friends and flat affect…"). Endicott and colleagues (1976) stated that the "basic idea and structure" of HSRS was retained, although the modified scale was renamed the Global Assessment Scale (GAS; p. 766).

In the DSM-III (American Psychiatric Association, 1980), the 100-point HSRS scale was transformed into a 7-point scale ranging from the highest possible score of 1 (superior) to the lowest possible rating of 7 (grossly impaired), with an option to rate as 0 (unspecified) if information was insufficient to determine a score (Bodlund, Kullgren, Ekselius, Lindström, & Knorring 1994). This scale was called "Highest Level of Adaptive Functioning Past Year" and was included as Axis V of the newly multi-axial DSM (discussed further below). The DSM-III provided a conceptualization of adaptive functioning as "a composite of three major areas: social relations, occupational functioning, and use of leisure time," although raters were instructed that "…there is evidence that social relations should be given greater weight because of their particularly great prognostic significance" (p. 28). In the DSM-III-R (American Psychiatric Association, 1987), the 7-point global function rating that served as the fifth axis of the DSM was replaced with a new measure titled the Global Assessment of Functioning Scale (GAF). This new measure represented a simple revision of Endicott et al.'s (1976) GAS, with a scoring modification such that scores ranged from 1 to 90 (with an option to score 0 for inadequate information). The GAF was conceptualized as an "overall judgment of a person's psychological, social, and occupational functioning" (p. 20), and raters were instructed to assign ratings for two time periods: current level of functioning at time of assessment and highest level of functioning in the past year. In the DSM-IV (American Psychiatric Association, 1994), the GAF scoring was changed to range from 1 to 100 (retaining the option to score 0 for inadequate information; see Table 3). The instructions were updated to state that clinicians could specify a time range for the rating (e.g., "current," "highest level in the past year"). The instructions specified that in certain settings, it may be more useful to assess social and occupational disability separate from symptomatology, and as such, the experimental Social and Occupational Functioning Assessment Scale (SOFAS) was included as an appendix. In the DSM-IV-TR (American Psychiatric Association, 2000), the GAF was retained as Axis V, although explanatory text in the DSM-IV-TR was added to clarify that the lower of symptomatic and functioning ratings should be assigned in instances in which there was a disparity, and that the lowest level of functioning over the past week should be considered when determining a ranking. However, clinicians were also given an option to specifically include a time period to which the score applies. The most recent version, DSM-5 (American Psychiatric Association, 2013a, b), eliminated the multi-axial system and removed the GAF entirely, for reasons discussed below. Despite its removal from the DSM-5, the GAF is still the legally mandated assessment for WC psychiatric injury PD evaluations.

Table 3 Example GAF scale values and associated WPI

GAF score	Summary	WPI
91–100	• Functioning is classified as superior; no symptoms	• 0
61–70	• Some functioning or relational difficulties but overall functioning is classified as pretty good; symptoms are present but they are mild	• 0–14
1–10	• Suicidal or homicidal threats or attempts; inability to maintain own hygiene	• 84–90
0	• Inadequate information	

When one considers that the role of the GAF in the DSM was intended to be just one part of a larger evaluative process, the extrapolation of this score as a stand-alone measure of functioning is arguably problematic. Specifically, the GAF was the fifth axis of a multi-axial diagnostic system. While Axes I and II were concerned with specific diagnoses, Axis III considered medical conditions that may be of relevance to an individual's psychiatric condition (e.g., a respiratory or digestive system injury), and Axis IV considered psychosocial and environmental contributors to an individual's diagnosis, treatment, or prognosis (e.g., financial problems, legal problems). Thus, Axis V was intended to provide, via the GAF, a global indicator that took into consideration the information from the other four Axes, providing an overall index of the individual's symptomatology and social and occupational general functioning.

An important question to consider concerns the purported role of the GAF and what information it is ideally meant to convey. The 7-point scale first introduced in the DSM-III was conceptualized as a measure of adaptive functioning, while the subsequent GAF was conceptualized as providing a composite assessment of psychological, social, and occupational functioning (Pendersen et al., 2007). Pendersen and colleagues (2007) have argued that the GAF usually serves two primary aims: (1) to demonstrate a need for psychiatric treatment and (2) to assess treatment outcomes. While a global

measure may suffice as a parsimonious measure of functioning in less high-stakes contexts in which clinicians or researchers require quick evaluations of overall functioning, the appropriateness of using such a measure to determine psychiatric disability as a means of assigning PD in injured workers is less clear.

More robust psychometric work on the GAF is necessitated to justify its use in this context. Without establishing that a GAF score can be validly interpreted as a sufficient sole indicator of PD (a highly unlikely finding), psychologists are essentially using an unsound measure to make extremely high-stakes decisions. This is all the more true when one considers what the GAF score replaced—the aforementioned eight categories of functional manifestation scores based on clinical observations and objective corroborating data for specified work functions. While a paramount issue with the prior measurement of disability was that the scale to evaluate the eight work functions was arguably subjective (ranging from minimal to severe), to simply abrogate this assessment and replace it, without scientific justification, with a single overall functioning score that requires perhaps greater subjectivity in scoring is problematic. To demonstrate that this level of information could be subsumed or improved by a single global score of functioning requires empirical data from psychometric validation studies.

The Psychometrics of the GAF

In a prescient observation, Piersma and Boes (1997) noted that, given its inclusion in the DSM, "It's ironic that the GAF has received so little formal research evaluation…it is likely that for many organizations…the GAF will become, de facto, a measure used to demonstrate clinical change…" (p. 36). Psychometric evaluations of the GAF remain limited in the years since this statement; the request for an "urgent focus" (p. 40) for improvement in the GAF raised by Piersma and Boes, given concerns about reliability and validity, has largely been ignored. While a full review of the psychometric properties of the GAF is out of the scope of the present paper (see Aas, 2010, for a systemic review of the GAF), it is useful to briefly describe the types of reliability and validity frequently employed in validation studies that would be relevant to the GAF in order to evaluate whether its use as a sufficient, single indicator of psychiatric PD for WC determinations is psychometrically (and subsequently ethically) justified.

Reliability

Reliability refers to the stability, consistency, predictability, and accuracy of scores for a given measure with the overarching goal of estimating the level of variance in a test that is attributed to error (Geisinger, 2013; Groth-Marnat, 2000). In relation to the GAF score, the two types of reliability of most relevance are inter-rater reliability and test-retest reliability.

Inter-rater reliability refers to the consistency in scoring across two or more independent raters (Cicchetti, 1994; Groth-Marnat, 2000). A seemingly basic but, in the case of the GAF, critical prerequisite to inter-rater reliability is that raters hold cognate understandings of the instructions. The instructions for the GAF advise the clinician to take into consideration current psychiatric symptoms (e.g., depression, sleep impairment) and social and occupational functioning, and synthesize the symptoms and functioning into a single GAF score. Specifically, the language in the DSM-IV-TR instructs clinicians to consider "psychological, social, and occupational functioning on a hypothetical continuum of mental health-illness" (American Psychiatric Association, 2000, p. 34). Given that it is unlikely that an individual will be equally impaired across the aforementioned three domains, the rater is instructed to consider the lowest score for symptomatology or functioning across the three areas and use this as the overall score. This combining of psychiatric symptoms and social and occupational functioning is not empirically grounded and has been questioned (Bacon, Collins, & Plake, 2002; Goldman, Skodol, & Lave 1992; Pendersen et al., 2007). Thus, per the GAF's instructions, the overall score—from which WPI is calculated and PDR percentage is granted—could theoretically be based solely on an individual's low social functioning alone, even if psychological and occupational functioning are in a high functioning range that would preclude PD. This is especially important in the WC context because scores of 70 and higher are directly associated with a 0 WPI score—typically with no room for further interpretation or explanation; thus, reliability in assigning scores even within a few points is important to ensure.

Studies exploring the inter-rater reliability of GAF scores have yielded mixed findings. Smith and colleagues (2011) described a discordant literature demonstrating both low (e.g., Bates, Lyons, & Shaw, 2002; Rey, Starling, Wever, Dossetor, & Plapp 1995) and high (e.g., Hilsenroth et al., 2000; Soderberg, Tungstrom, & Armelius, 2005) inter-rater reliabilities. Loevdahl and Friis (1996) reported that inter-rater reliability for the GAF can be high among experienced raters but that inter-rater reliability among untrained raters can fall in the unsatisfactory range. Among the 104 raters evaluated, systematic deviations of −23 to +23 points were identified; 80 % of raters demonstrated rater bias of $\geq \pm 11$ points.

Similarly, in a 2007 review exploring the disparate inter-rater reliability findings, Vatnaland and colleagues found that while a cursory examination of GAF inter-rater reliability via interclass correlations (ICC) suggests excellent (i.e., ICC> 0.74) reliability, methodological shortcomings in the majority of the reviewed studies rendered the results problematic. Specifically, the authors found that many of the studies exploring inter-rater reliability of GAF scores employed

"conditions highly unrealistic" (p. 327). For example, they found extensive prior training and calibration, clinician samples coming from single research settings (i.e., "within-center inter-rater reliability" p. 327), and unblinded methodologies such that researchers were aware that their GAF scores were being evaluated. In their own study of GAF ratings in a realistic, acute psychiatric context, Vatnaland et al. (2007) compared ICC coefficients between scores obtained in an acute psychiatric hospital setting (i.e., closer to the real world) and in a research setting and found them to be only 0.39 and 0.39 at admission and 0.56 and 0.59 at discharge (note that an ICC between 0.40 and .60 is considered *fair*). There is no formal training available for the GAF for QMEs in WC settings and no system of checks and balances to attempt to ensure inter-rater reliability.

Given the importance of the unilateral evaluator rating on the GAF provided by the QME in determining PDR, "whether the GAF scale can be properly used in a variety of contexts and by different raters, and still display high inter-rater reliability" (Vatnaland et al. 2007, p. 326) represents a key consideration. The parties involved in a given case should feel confident that the rating given to a worker alleging injury will not significantly vary depending on to which evaluator they are assigned. A large-scale study with a random sample of QMEs of various levels of expertise (i.e., years licensed, years working in WC settings, board certifications) comparing GAF ratings for vignettes of varying complexity is an example of the type of study that would be important to undertake in order to strengthen confidence in the reliability of the GAF. Loevdahl and Friis (1996) recommended that, in real-world settings, procedures be in place to identify raters with extreme deviations via random quality assurance checks where experienced raters also provide ratings of the same individuals and scores are compared so that further guidance around scoring instructions can be provided to raters.

The second type of reliability that would be important to explore in WC contexts is test-retest reliability. Test-retest reliability, also called the coefficient of stability, captures the temporal stability of scores (Geisinger, 2013). It is important to first have a firm grasp of the nomological net—to know exactly what one hopes to measure in the context of functioning—in order to make reasonable hypotheses about the degree of temporal stability, and thus level of reliability coefficient, that one would expect (Cicchetti, 1994; Groth-Marnat, 2009). In WC, the PDR is meant to occur after the period of maximum recovery has been reached, and thus the temporal stability of the GAF should arguably remain fairly constant if the same individual is re-assessed by the same evaluator at multiple time points following the PD assessment. In one of the few studies on the test-retest reliability of the measure, conducted in PTSD populations, the test-retest reliability was modest (Miller et al., 2008). The dearth of more generalizable test-retest reliability studies of the GAF has been called a gap in knowledge (Aas, 2010). Longitudinal studies are needed in order to demonstrate whether scores behave as expected (i.e., that patients granted PD remain in a range of scores consistent with disability status).

Aas (2010) suggested that a primary hindrance to the reliability of the GAF may be that scoring instructions are not intuitive for evaluators. The vagueness of the instructions of the GAF may be the largest hindrance to the inter-reliability of the measure. For example, the instructions ask clinicians to exclude physical limitations in determining the global score: "Do not include impairment in functioning due to physical (or environmental) limitations." However, the way that evaluators make scoring decisions is an empirical question that has received little attention. There have been attempts to modify GAF scoring in order to bolster inter-rater reliability (see Hall, 1995 and Kennedy, 2003); however, no approach has been widely adopted. Bacon, Collins, and Plake (2002), in a study asking three raters to provide reasons for GAF scores in a research setting, found that GAF ratings were strongly based on decisions *other than* adaptive functioning/impairment (e.g., symptom severity). They concluded that "the GAF is not a good measure of adaptive functioning, yet important decisions affecting clinicians and clients are made on the basis of GAF scores" (p. 202). Two interesting, and to the authors' knowledge unexplored, areas of inquiry would be (1) the types of decision-making that clinicians working in WC settings utilize in disentangling physical limitations in global functioning, especially in the presence of physical injuries, and (2) whether this stipulation poses a valid counterargument to one of the key rationales for limiting physical-mental claims (i.e., that the mental disability has been taken into account when assigning a PD rating for the physical injury, whereby including a separate mental claim would be in essence "double-counting" the same injury).

Validity

Validity refers to an overall evaluation of the adequacy of the evidence to support the appropriateness of using the scores on a measure to draw inferences and interpretations about a given construct (Messick, 1995). Reliability is a necessary, but not sufficient, requisite for validity, the latter of which can be separated into several categories, the most relevant of which for the current purposes is criterion validity (i.e., concurrent, predictive, and discriminative). Although these issues will be briefly covered below, it is important to remember that, because validity is constrained by reliability, and given the aforementioned problems with reliability, validity is by definition problematic. Validity not only describes whether a measure does indeed evaluate what it purports to assess but also provides insights into the meaning of the scores (Cicchetti, 1994; Foster & Cone, 1995; Reise, Waller, & Comrey, 2000;

Bornstein 2011) Building upon Foster and Cone's (1995) recommendations, two important considerations that should guide any psychometric evaluation of the GAF are that (1) the GAF score is assumed to assess the degree to which workers who allege psychiatric injury have psychiatric symptoms and functioning deficits and (2) the purpose of the GAF score is to categorize workers in terms of disability. The latter is an especially important consideration in WC settings given that scores of 70 and higher preclude PD compensation (i.e., they are associated with a WPI of 0) and the actual compensation amount for individuals with functioning scores below 70 is directly calculated from their GAF score (e.g., a GAF score of 65 is given a WPI of 8; a GAF score of 35 is given a WPI of 61).

Criterion validity is concerned with the relationship between scores on a given measure and scores on criteria to which an instrument should be practically related (Cicchetti, 1994; Foster & Cone, 1995; Groth-Marnat, 2009). Criterion validity is typically described as being concurrent, predictive, or discriminative. In a review, Aas (2010) described shortcomings in the validity of the GAF and suggested that problems with concurrent and predictive validity were at the core of the problem. Concurrent criterion validity is concerned with correlations between the scores on the measure being evaluated and scores on existing measures of a relevant criterion; the scores on the measure are typically compared with the scores on a "gold standard" measure in the field. While evaluation of criterion validity is fairly straightforward for many constructs that have been more clearly operationalized (e.g., depression), this is a much more difficult evaluation for measures of less clearly operationalized psychological phenomenon, such as global functioning. Further adding to the complexity, in the WC setting, the global functioning measure is used as a proxy for disability. This is an important consideration in the context of criterion validity because it is unclear as to what the "gold standard" criterion should be (e.g., disability in specific domains of functioning; inability to complete work functions).

The limited literature summarizing concurrent validity for the GAF is discordant, due in part to the lack of clarity around with what one would hope that the GAF would be correlated and in part because of the dearth of rigorous psychometric studies. While Aas (2010) listed several studies reporting problematic criterion validity, Burlingame and colleagues (2005) described moderate to high concurrent validity in the GAF, citing Hilsenroth and colleagues (2000) and Startup, Jackson, and Bendix (2002). However, both of these studies are limited in their external validity, and neither of these studies used disability as the criterion. Specifically, the Hilsenroth et al. (2000) study found that GAF scores were significantly related to concurrent patient scores on the SCL-90-R global severity index (thus here, symptom severity was the purported criterion). This was in a sample of 44 patients admitted to an outpatient university clinic in which assessments were given by clinical psychology doctoral students who were trained in the assessments prior to study commencement. The GAF was not given as a stand-alone measure of functioning, but along with two other functioning scales available in the appendix of the DSM-IV that specifically assess social and occupational functioning and relational functioning. Thus, the modest sample size, having raters who were aware of the research purposes of their ratings and who underwent group trainings on GAF scoring, and the presentation of the GAF along with *separate* ratings of social and occupational functioning and relational functioning (i.e., the raters knew that these facets of functioning would be captured in other ratings) arguably limit the external validity of the study. The Startup et al. study concluded that correlations between the GAF and other measures of symptoms and functioning were "large and highly significant" (p. 421); however, they acknowledged that this was only true in follow-up assessments and not in the intake. Further, the sample consisted of $N=64$ patients with a diagnosis of schizophrenia suffering an acute psychotic episode and criterion measures were schizophrenia-specific (e.g., Scale for the Assessment of Negative Symptoms), also limiting generalizability of the findings.

In terms of criterion validity in WC settings, the GAF, which is used as a proxy of PD and from which a WPI score is calculated, should arguably be associated with the ability to complete work functions and with disability in specific domains of functioning (e.g., mobility). The use of the GAF in WC settings implies that the criterion is PD that requires compensation because the individual cannot complete specific functions. However, there is a dearth of studies evaluating the relationship between GAF scores and the criterion of disability. In one of the few available studies, Parker and colleagues (2002) administered the GAF as part of a battery of disability and functioning measures to a small sample of patients ($N=69$) with a serious psychiatric diagnosis and found that the GAF was only moderately associated with the other measures; of interest, the authors noted that the GAF was not a good measure of disability. Aas (2010) described that any measure of functioning should be evaluated as to (1) which types of functioning should be assessed, (2) how to grade the various types of functioning, and (3) whether a single, global score is sufficient as an aggregate index of the types of functioning being assessed. Aas further noted that, despite numerous international efforts to develop rigorous measures of functioning, this research has not informed the development or updates of the GAF.

Aas (2010) also described the predictive validity of the GAF as problematic (see Hay et al., 2003, Parker et al., 2002, among others). Conceptually, as the GAF score that determines PD compensation should be given after MMI, the scores should remain consistent (i.e., a patient who was in a given range of disability at year 1 should remain in that

range at year 2). Predictive criterion validity is concerned with predicting scores on related measures that will be administered at a future time point or that enable inferences to be made about another criterion, particularly in terms of monitoring change. Moos, McCoy, and Moos (2000) assessed predictive validity of the GAF in a sample of 1688 patients with substance use disorders and comorbid psychiatric disorders. The authors reported "little if any relationship between ratings of patients' current or highest level of global functioning and psychological, social, or occupational functioning at 1 year follow-up" (p. 458). Although this study was in the context of substance abuse and in a Veteran's Administration (VA) setting, factors which can limit the external validity, the GAF ratings were assigned as part of routine clinical diagnostic interviews (i.e., not under controlled research conditions), and thus this aspect of the study is closer to the real-world settings in which the GAF score is typically assessed.

Discriminative validity is another subtype of criterion validity that refers to a measure's ability to discriminate between known groups. A priori hypotheses about the known groups must be posed; otherwise, observed differences should be attributed to potential test bias and not as evidence of discriminative validity. In a large sample (N=283,754) of both inpatient and outpatient Veterans using existing VA data, Greenberg and Rosenheck (2005) found that indicators of greater severity in mental illness (e.g., schizophrenia diagnosis), disability ratings of 50 % of greater, and inpatient versus outpatient status were all associated with lower GAF scores. The authors concluded that the results provided support for the discriminative validity of the GAF scores. However, in WC settings, the necessary discriminative abilities of the GAF are more specific; because scores of 70 and higher on the GAF are associated with 0 WPI, scores must discriminate, with optimal sensitivity and specificity, between psychiatrically disabled and non-disabled workers. However, to date, there has been no empirical evidence provided to justify the cutoff of 70. In the same vein, validation studies need to be undertaken in WC settings demonstrating a qualitative difference between workers scoring 70 and higher (i.e., that they are not disabled and can work) and those scoring below 70. This is all the more salient when one considers the continuous scale of the GAF; if reliability is strong, continuous (versus categorical) scaling can theoretically be a good quality for detecting subtle differences in scores (i.e., sensitivity; Aas, 2010). However, without empirical studies demonstrating the relationships between specific numerical GAF scores and disability, a continuous scale with individual numerical scoring options suggests a sensitivity that may not be appropriate. Aas describes a need for "statistically significant differences for samples with small differences in the severity of the symptoms" (p. 22).

A global measure that collapses various domains of functioning inherently implies the similitude and covariance of constructs that may, in fact, present quite differently (First & Pincus, 2002). In a study of GAF ratings in the context of mental health-care outcomes and treatment allocation in VA substance use and psychiatric disorder settings, Moos, Nichol, and Moos (2002) found that symptoms and clinical diagnoses were stronger predictors of veterans' GAF scores than were social or occupational functioning. Others have also questioned to what level of functioning is taken into account when raters are determining GAF scores (Bacon, Collins, & Plake, 2002). If studies undertaken in WC contexts were to also show that scores are primarily driven by symptoms and not functioning, it would suggest that the GAF would be more accurately presented as a measure of psychiatric symptomatology than social or occupational functioning. This would seriously call into question the appropriateness of using the measure for its present purposes of making PD decisions. Even assuming enhanced training among raters that would result in more systematic scoring processes (i.e., higher inter-rater reliability), there is the larger question of whether a global score provides useful information given the lack of empirical support for the reliable covariance between symptoms and functioning, an assumption on which the GAF is grounded (Narrow & Regier, 2013). While this may not be a central concern for certain evaluative contexts, reasons for assuming covariance of psychiatric symptoms, occupational functioning, and social functioning for the purposes of determining the extent of psychiatric disability in WC settings should be addressed and empirically studied.

APA Ethics Code

In a 1992 exploration into the role of psychological testing in forensic assessment, Heilbrun wrote that "The appropriate role of psychological testing in forensic assessment has been debated for years and is far from clear at present." (p. 257). Contributions to this interesting ethical space that lies at the interface of law and psychology, particularly in the field of WC, have not progressed much since this initial statement. The issues highlighted below are important considerations for psychologists involved in WC settings given that such individuals may be placing themselves in ethically precarious situations per the APA enforceable standards while abiding with their legal obligations and evaluative roles.

It is useful at this stage to turn to the APA Ethics Code (hereafter Ethics Code; American Psychological Association, 2010) in order to better understand the ethical problems associated with using an unvalidated measure in the WC context, as the Ethics Code for the field mandates appropriate validation. The structure of the Ethics Code includes five general principles and ten ethical standards. The general principles are described as "aspirational in nature" and are meant to promote the value of ethical excellence for the field, and thus they are not stipulatory (e.g., *Principle E:*

Respect for People's Rights and Dignity). In contrast, the ethical standards are enforceable, with deviations constituting cause for sanctions (e.g., *Standard 9.02: Use of Assessments*). Given the seminal role of assessments in psychology—and their great potential for misuse—there is an entire standard dedicated to ethical considerations in the realm of assessments. The following section will briefly consider Standard 9, which is devoted to assessments, in the context of how PD is currently evaluated by psychologists in the California WC system. The following section will provide a summary of some of the key ethical problems worthy of further exploration (see Table 4 for a summary of the points made below).

Commentary of Standards 9.01–9.10 in the Context of WC PD Evaluation

Standard 9.01: Bases for Assessments Standard 9.01 specifically mentions forensic testimony and report writing, dictating that evaluations should be based on "information and techniques sufficient to substantiate…findings." Because the GAF is essentially the sole assessment utilized to quantify permanent psychiatric disability in evaluations, it becomes the *sole* technique used to "substantiate" findings regarding disability in this context. While this practice is sufficient for meeting the qualifications of the law, given the aforementioned psychometric issues with the GAF, it is questionable whether this technique is sufficient ethically as no one instrument is typically considered "sufficient" to make a psychiatric determination or diagnosis (Groth-Marnat, 2009).

Further, the standard recommends that psychologists only provide their opinions *after* conducting an "adequate" examination. INDUSTRIAL MEDICAL COUNCIL, PSYCHIATRIC PROTOCOLS (1992) (amended 1993) outlines the information clinicians must gather in order to write a medical report following the evaluation in a effort to standardize report writing across QMEs, noting however that the guide is "more suggestive than prescriptive" (p. 2). Further, psychological testing is described as "an additional source of information," (p. 8) implying that it is not required in the report. Thus, presumably, the psychologist could write the evaluative report without conducting a single standardized assessment apart from the GAF, per the 2005 updated PD schedule, and be in line with what is required in the forensic setting while arguably not meeting ethical requirements.

Additionally, the issue of utilizing records and past medical history, which is often done in such evaluative contexts in

Table 4 Summary of ethical concerns regarding use of the GAF in WC PD evaluation settings per Standard 9 of the ethics code

Standard	Areas of concern
9.01: Bases for assessments	• Using the GAF without a clear justification in WC determination • Parsimony of GAF can preclude transparency in scoring practices • Potential for "cherry-picking" of medical records
9.02: Use of assessments	• Lack of psychometric robustness and problematic extension of the GAF beyond its intended purpose • Limited information about cultural validity of the GAF • No guidance on how to proceed in cases with suspected malingering or exaggeration of symptoms
9.03: Informed consent in assessments	• HIPAA restrictions • Lack of standardization in what is required in informed consent in WC evaluations • Limits to confidentiality • Lack of standardization in how to explain nature and purpose of each assessment offered to injured workers • Question identification of the client and how to address this in obtaining consent
9.04: Release of test data	• Lack of guidelines regarding how to address release of raw and/or scaled scores to injured workers, third parties, payer of services, etc. • Problems with reliance on psychologists' scoring in assessments with subjective criteria (e.g., GAF) and future replicability
9.05: Test construction	• Insufficient nomological net (i.e., what construct is the GAF meant to be measuring in this context?)
9.06: Interpreting assessment results	• Potential for over-pathologizing various cultural contexts, including that of "the injured worker"
9.07: Assessment by unqualified persons	• Lack of training requirements for use of the GAF
9.08: Obsolete test and outdated test results	• Continued use of the GAF and DSM-IV-TR criteria despite its removal from the DSM-V • Need for oversight by professionals with mental health expertise to update WC evaluation guidelines consistent with new evidence
9.09: Test scoring and interpretation services	• Boundaries of competence in GAF scoring • Burden of responsibility for the appropriate application of assessments rests on the psychologist
9.10: Explaining assessment results 9.11: Maintaining test security	• Insufficient guidelines regarding the role of psychologists in explaining the results to the injured worker and to all other individuals who will have access to the report • Insufficient guidelines regarding what materials must be released and to whom when supplementary assessments (e.g., MMPI, BDI) have been used to inform a GAF score

determining a GAF rating, is referenced in 9.01(c) of the Ethics Code as permissible provided that the sources of information are transparent. However, this too can be ethically precarious. Schatman and Thoman (2014) explore the problem of "cherry-picking" records and note that the bottleneck to providing a full history is likely to come from claims managers in such cases. They recommend that psychologists acting as medical examiners not blindly accept the records with which they are provided as complete. Rather, in order to protect the integrity of the case, psychological medical examiners should keep in line with their own ethical duties, to "…take preemptive measures to ensure…access to *all* of the relevant records prior to conducting an examination" (p. 195). While Schatman and Thoman cited aspirational Principle D (fairness and justice) in support of this argument, the present authors would add that per Standard 9.01, negligence in this area may actually constitute an enforceable violation.

Finally, in terms of other professionals attempting to access or reference the psychologist's PD report, it would arguably be challenging to understand the thought processes behind assigning a GAF score (e.g., how potential disparities in symptoms versus social and occupational functioning were weighted; ensuring that physical and environmental limitations did not factor into the score per the GAF's instructions, etc.). Without commentary elucidating the single numerical rating, the rationale behind assignment of a specific score may not always be comprehensible (and technically is not required).

Standard 9.02: Use of Assessments Standard 9.02 stipulates that psychologists utilize (i.e., administer, score, interpret) assessments supported by up-to-date research on evidence and utility for the given assessment. Further, assessments should demonstrate reliability and validity established specifically for the population with which the test is being used and for people with varying sociodemographic (e.g., gender, ethnicity) characteristics who might complete the measure. Thus, it is incumbent on psychologists using assessments in WC settings to use measures with proven psychometric validity, not just generally, but specifically for the population being tested. Typically, this same rigor is not required in guidelines written for psychologists serving as QMEs in WC settings. Thus, a psychologist could conceivably write a report satisfying all of the requirements for an appropriate QME evaluative report for alleged psychiatric injury that is psychometrically unsupported and thus ethically unsound per the APA Ethics Code. The sole *required* assessment is the GAF, emphasizing the salience of this single score.It is particularly within the context of Standard 9.02 that the various psychometric limitations of the GAF, described in the prior section, emerge as highly concerning. Questions regarding the validity and reliability of the measure's score preclude its confident use by psychologists, especially for such high-stakes assessments. Moreover, concerns regarding the psychometric rigor of the assessment for the specific population in question, in this case allegedly injured workers, have not been addressed in the literature; further, the measure must be a reliable and valid measure for all workers regardless of sociodemographic characteristics. A primary concern with this population is malingering (Greve, Bianchini, & Brewer, 2013), but given that this is a single-item assessment, there is no malingering/dissimulation scale associated with the measure. Thus, it is unclear how a psychologist who suspects exaggeration of symptoms or reported functioning deficits should proceed. Technically, a score of "0" can be granted in cases lacking sufficient evidence, and this may be the appropriate course of action in such instances. However, there is a lack of guidance regarding this highly probable scenario.

Given the salience of the GAF in evaluating psychiatric disability and assigning impairment, more rigorous psychometric validation of the GAF score in this context is recommended. To the authors' knowledge, no psychometric validation studies exploring the reliability and validity of the GAF in injured worker populations have been published, which begs for increased psychometric rigor in this area.

Standard 9.03: Informed Consent in Assessments Standard 9.03 requires that psychologists obtain informed consent for most assessment activities. In addition to standard confidentiality concerns, psychologists working with injured workers should arguably underscore and clarify the purposes of the current assessment, to whom the results will be provided, and limits to confidentiality. In an effort to navigate the complex waters of health-care record release in the age of the Health Insurance Portability and Accountability Act (HIPAA), Borkosky, Pellet, and Thomas (2014) argued that, while many argue that, "HIPAA does not regulate forensics", making reference to a 2003 Connell and Koocher argument precluding forensic evaluations from HIPAA disclosure rules, such protections are rarely upheld. In practice, mental health practitioners often must release data per Ethics Code Standard 4.05 (Disclosures) and Standard 9.04 (Release of Test Data). Evaluators are in fact specifically permitted to release records from WC evaluations [45 C.F.R. § 164.512(l) (2011); Borkosky et al., 2014]. While some psychologists working in forensic settings include a description of HIPAA and a clear explanation of to whom data may be released, this is neither standard nor standardized. Thus, injured workers may be participating in evaluative activities without fully understanding the ramifications of how the scores (i.e., GAF rating) will be used.

Questions regarding to whom the information disclosed will be used and purposes of the assessment may influence how forthcoming an injured worker is with information granted. Accordingly, systematizing the consenting process

or, at a minimum, requiring standard language regarding the meaning of the GAF, what the score means, and how it is being determined and clarifying to whom scores will be released can help demystify the process. Although the APA Standards may not address the issue directly, "transparency" is moving toward consideration of ethical conduct by mental health-care providers more generally (Kahn, Bell, Walker, & Delbanco, 2014). Finally, the question of "Who is the client?" is a common one in forensic settings, especially in contexts such as WC where they may be different parties paying for the evaluation (e.g., In the case of QMEs, is the state of California the client, or the injured worker on whom the evaluation is being conducted considered "the client"? In the case of an MPN treating an admitted claim, is the client the insurance company or the individual being evaluated?), and this too should be carefully considered by the evaluating psychologist and potentially addressed in the consenting process.

Standard 9.04: Release of Test Data The issue of whether to release the raw data, rather than aggregate findings, to patients or other relevant individuals is a contested issue in psychology. In the case of the GAF, the score is a single numerical value; while the release of this single score may not be problematic, the score is presumably made via the clinician's professional opinion based upon data such as the clinical interview, medical records, and any other utilized assessments. It is unclear regarding what exactly constitutes "test data." Per Standard 9.04, test data specifically refers to "raw and scaled scores, client/patient responses to test questions or stimuli, and psychologists' notes and recordings concerning client/patient statements and behavior during an examination." Concerns around the potential misuse of raw scores by those not trained in the interpretation of results and uses of the scores beyond those for which they were intended make some psychologists wary of their release. Further, in the event that a GAF score is contested, the question of what supporting materials suffice to justify the score and to whom they could and should be provided is nowhere addressed.

Standard 9.05: Test Construction In essence, all the parties involved in a legal case are asked to have full faith in the construct validity of the legally mandated tests (i.e., that tests are measuring the constructs that they purport to measure; Groth-Marnat, 2009). The developers of the GAS, the predecessor to the GAF, cautioned that ratings should be based only on functioning for a specified time period and should not be influenced by "considerations of prognosis, previous diagnosis, or the presumed nature of the underlying disorder" (Endicott et al., 1976, p. 767). It is important to consider the original intent of the developers of the GAF and whether this is consistent with its current use in WC settings. Best practices in measure development dictate that the construct of interest must first be carefully defined via a nomological net (Cronbach & Meehl, 1955). That the term "nomological" is derived from the Greek for "lawful" is apt given that one can conceptualize the nomological network as the core evidence for what facets of a given construct (e.g., functioning) an assessment instrument or technique should capture (Bornstein 2011) A complete nomological net should include the internal structure of the construct, a map of relationships among the construct of interest and other related constructs that make explicit not only to what the construct relates but also to what the construct does not relate, and a clarification of the target population for whom the measure is intended (Clark & Watson, 1995). Phelan, Wykes, and Goldman (1996) describe the challenge of such in the context of global function scales because "Functioning is an abstract concept, incorporating a range of abilities…Because the notion of global functioning covers all these abilities there is little consensus about the precise meaning of the term." (p. 15). Indeed, a primary impetus driving Dimsdale and colleagues' (2010) advocacy for the replacement of the GAF with a five-dimensional psychiatric Apgar was precisely the lack of well-defined "conceptual anchors" (p. 515).

Standard 9.06: Interpreting Assessment Results Standard 9.06 serves to ensure that when appropriate, scores are not reported and interpreted in a vacuum, but within the context of the assessee's cultural context and "situational, personal, linguistic, and cultural differences" that may limit the psychologist's interpretation of results. In describing the GAF, Dimsdale et al. (2010) noted that "Implicit in the scale are assumptions about what constitutes mental health and mental illness and what constitutes well-being in society" (p. 515). In perusing the various examples that the GAF provides for each level, there are potentially problematic items. For example, an example in the 41–50 range includes "severe obsessional rituals," but high levels of religiosity can erroneously be diagnosed by clinicians as obsessions (Allmon, 2013). Additionally, even in taking a broad view of culture and considering the "cultural context" of injured workers, many of the work-related examples that the GAF provides in attempting to operationalize the various rating categories (e.g., "unable to keep a job" in the 41–50 range; "depressed…and is unable to work" in the 31–40 range) are problematic in light of the fact that the population of interest is workers seeking compensation and being evaluated for PD, which may lead to unnecessary pathologizing and inaccurately low GAF scores.

Standard 9.07: Assessment by Unqualified Persons Standard 9.07 is an effort to limit the use of psychological assessments to those psychologists trained to administer the specific assessments. The law requires that any information beyond "purely clerical intake data" is collected by a doctoral-level-trained psychiatrist or psychologist [DIVISION OF WORKERS'

COMPENSATION-MEDICAL UNIT: APPLICATION FOR APPOINTMENT AS QUALIFIED MEDICAL EVALUATOR (2013)]. Thus, the underlying assumption is that a doctoral-level psychologist should be proficient in assessing an individual in order to determine an overall functioning GAF score. Simply having generic doctoral-level training in psychology does not necessarily mean that one has had training in the GAF. However, given the aforementioned research demonstrating that training may improve inter-rater reliability of the GAF and the seminal role of the GAF in PDR for alleged psychiatric disabilities, psychologists in WC evaluative contexts should arguably be required to seek specific training in the GAF prior to undertaking assessment duties.

Standard 9.08: Obsolete Tests and Outdated Test Results Standard 9.08 provides a timely context for consideration of the current use of the GAF, given that the GAF was removed from the DSM-5. The California Labor Code states that "A psychiatric injury shall be…diagnosed…using the terminology and criteria of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Third Edition-Revised, or the terminology and diagnostic criteria of other psychiatric diagnostic manuals generally approved and accepted nationally by practitioners in the field of psychiatric medicine." [see CAL. LAB. CODE § 3208.3 (West 2011)]. When the DSM-IV and DSM IV-TR became available in 1994 and 2000, respectively, the WC evaluative community transitioned to using these diagnostic manuals. However, given the substantial changes to the DSM-5 (e.g., the removal of the GAF, the elimination of a multi-axial system), coupled with the HIPAA requirement for World Health Organization International Classification of Disease (ICD) diagnoses that commenced in October 2014, the direction toward which various WC systems will turn remains unclear. In an official response to clinicians addressing implementation queries, the American Psychiatric Association answered the question of how disability and functioning should be assessed given that the GAF is no longer part of the DSM-V, stating "We do not believe that a single score from a global assessment, such as the GAF, conveys information to adequately assess each of these components, which are likely to vary independently over time," and instead recommended supplementing clinician evaluations of suicidal and homicidal behavior risk, symptom severity, and diagnostic severity with the disability rating the World Health Organization Disability Assessment Schedule (WHODAS 2.0) stipulated "for those who relied on a GAF number" (APA, 2013b). The response deemed the WHODAS 2.0 (discussed further below) the "best current measure of disability for routine clinical use" by the DSM-5 Disability Study Group (APA, 2013b, p. 2). However, the WHODAS 2.0 also has limitations (e.g., no normative values, threats to validity by virtue of being a self-report, reliability challenges in terms of detecting disability in individuals with premorbid high functioning) and is subject to the same primary criticism to which the GAF is subject by virtue of being a single, numerical indicator of global functioning (Gold, 2014).

Standard 9.09: Test Scoring and Interpretation Services While Standard 9.09 primarily refers to those instances wherein psychologists use automated or contracted scoring services, substandards b, which refers to boundaries of competence in scoring, and c, which refers to responsibility for the application, interpretation, and use of instruments, should be noted. Presumably, a psychologist could request consultation in determining a GAF score, but in such cases, the burden of responsibility would still fall on the primary evaluator. If psychologists serving in evaluative capacities do not feel comfortable assigning the GAF score, however, such consultation may be warranted. The reference to Standard 2.01 (Boundaries of Competence) for the interpretation of assessments is important because it emphasizes that the ethical duties of the psychologist do not end with simply assigning the GAF score. The ethically minded psychologist should be knowledgeable regarding how the GAF scores will be utilized and what purpose it will serve. Specifically within the context of WC, the GAF score is directly translated into an impairment rating and subsequent disability award, and thus psychologists should be mindful and aware of such. This is particularly relevant in instances in which there may be factors associated with a GAF score that may not fully represent the scope of disability for a given injured worker, either in terms of being inflated or deflated.

Standard 9.10: Explaining Assessment Results Standard 9.10 explains that while generally psychologists are expected to provide feedback to individuals following completion of testing to ensure that evaluees fully comprehend the results, there may be certain settings (e.g., forensic settings) in which doing so is precluded. While workers can presumably ask to know their GAF score, there are no standardized guidelines that define the role of psychologists in WC evaluative settings in terms of stipulated requirements or limitations in further explaining assessment results (e.g., what factors influenced a specific numerical rating) to the allegedly injured worker and to all other individuals who will have access to the report.

Standard 9.11: Maintaining Test Security Psychologists are required to "make reasonable efforts to maintain the integrity and security of test materials and other assessment techniques," with test materials referring to manuals, instruments, protocols, and test questions. This standard is designed to protect multiple entities, including the copyrights of test developers. Test materials are distinguished from "test data," referenced in Standard 9.04, and include

manuals, instruments, protocols, and test questions. Because the GAF is available as an open access item, there is little concern regarding its release. However, in the event that a psychologist used previous records, including test scores, or administered other measures (e.g., the MMPI) and used their scores to justify a GAF score, there could theoretically be requests from the multiple parties involved for the release of this information. Thus, there can be conflicts between efforts to respect the intellectual property of the test developer and desires on the part of the various parties to understand which specific items contributed to a given GAF score. For example, one can consider the case of assigning a GAF score when the lowest rating came from depressive symptomatology that the psychologist inferred from assessments in the patient's file measuring depression. This is an area of growing concern for psychologists acting in forensic roles because "courts increasingly order disclosure of test materials during discovery" (Kaufmann, 2009, p. 1131).

Recommendations for Increasing Psychometric Rigor and Revisiting Psychiatric Permanent Disability in California

As referenced earlier in the paper, existing RAND data regarding psychiatric injury in WC demonstrate that psychiatric impairment cases have among the highest earning losses (with average earning loss reported at 38 %; Reville et al., 2005). In fact, prior to the 2013 SB 863 elimination of tiered FEC adjustments, the aforementioned RAND finding spurred a change in how PD was adjusted. Per the January 2005 Schedule for Rating Permanent Disabilities, "… a psychiatric impairment receives a higher FEC adjustment because RAND data shows [sic] that a relatively high wage loss corresponds to the average psychiatric standard PD rating" [LABOR AND WORKFORCE DEVELOPMENT AGENCY, DEPARTMENT OF INDUSTRIAL RELATIONS, DIVISION OF WORKERS' COMPENSATION, SCHEDULE FOR RATING PERMANENT DISABILITIES UNDER THE PROVISIONS OF THE LABOR CODE OF THE STATE OF CALIFORNIA (2005)]. However, SB 863 (1) removed the condition-specific FEC and (2) severely limited the compensation for psychiatric injuries resulting in PD. This limitation was stipulated in response to perceived rampant abuses in the system. However, these abuses may be largely attributed to the subjective, flawed, and ethically dubious means of measuring psychiatric disability.

Unfortunately, rather than empirically exploring the psychometric quality of existing measurement of psychiatric disability (i.e., the GAF) such that the most sound assessment guidelines can be provided, California's limitation of compensability of many types of psychiatric disability (i.e., physical-mental claims) serves to delegitimize psychiatric disability without addressing the source of the problem. The progress that California has made toward utilizing empirical findings (e.g., from RAND Institute for Social Justice) and embracing more objective measures in revising physical disability rating and compensation should be paralleled in the state's approach to psychiatric disability as well.

There have been many proponents advocating for an alternative or supplementation to the GAF. For example, more than two decades ago, Goldman, Skodol, and Lave (1992) recommended that the GAF be used to reflect symptoms and psychological functioning alone and that two other distinct scales could supplement the GAF, assessing (1) relational functioning and (2) social and occupational functioning. The DSM-IV added a Global Assessment of Relational Functioning (GARF) Scale and the Social and Occupational Functioning Assessment Scale (SOFAS) in an appendix as experimental scales. However, even if administered, these are not used in WC PD determinations. Dimsdale et al. (2010) recently argued for a psychiatric Apgar-like scale that considers the five dimensions of neurocognitive functioning distress, psychiatric features, everyday functioning, and social relationships, with the first three dimensions obtained from a mental status exam and the latter two garnered from a detailed psychiatric and social history. Others have advocated for a "split GAF," such that there is one scale relating to symptomatology and another assessing social/occupational functioning (Pedersen, Hagtvet, & Karterud, 2007). Young (2008), one of the few forensic experts specifically addressing the GAF in the context of disability evaluations, recommended a revised GAF that is consistent with the DSM-IV and the AMA Guides assessments of global functioning; he suggests a reevaluation of the rating system and offers a suggestion for a revised scoring scheme that is based on the categories absent, mild, moderate, severe major, very serious, dangerous gross, and complete.

Another argument for rethinking the GAF comes from the American Psychiatric Association. In deciding to remove the GAF entirely from the DSM-V, the American Psychiatric Association recommended the use of the aforementioned WHODAS 2.0 for assessments of disability (Narrow & Regier, 2013). Unlike the GAF, the WHODAS 2.0 is a 36-item measure (although short forms are also available) that can be clinician- or self-administered, has been designed with consideration of cultural validity, and assesses functioning in the following six domains: cognition, mobility, self-care, getting along, life activities, and participation (Üstün et al., 2010). However, Gold (2014) argues that "No single number can convey enough information to address adequately all the different domains of functioning…" potentially impacted by psychiatric disorders (p. 180). Further, Gold references numerous psychometric shortcomings and gaps of the WHODAS 2.0 that preclude its use in forensic settings given that it is designed to be a self-report assessment; the reliability and validity of the clinician-rated versions have not been fully explored. A glaring threat to validity in forensic settings is that

the WHODAS 2.0 does not differentiate between impairment that is due to physical or psychiatric symptomatology, and thus medical and psychiatric impairment can be confounded (Gold, 2014). Thus, the WHODAS 2.0 would not be a sound substitution for the GAF.

The present authors recommend that efforts be undertaken toward developing an evidence-based assessment protocol that has clear evidence of psychometric strength for the intended purpose of determining WC disability. Such an effort calls for interdisciplinary collaboration that utilizes psychologists and researchers with expertise in psychometrics in the process. Until that time, psychologists serving evaluative functions should be aware that the current legally mandated method for evaluating PD in WC settings poses numerous ethical concerns. There is an urgent need for further guidance and updated evaluation mechanisms for psychologists. Until such guidance and updated methodology are available, some evaluators may choose to refuse to engage in the present, ethically compromising practices. Others may find it useful to supplement the required assessment (i.e., the GAF) with other validated assessments and to include language in the report that clearly states that the current method is not psychometrically advised but is being included because it is at present the only legally mandated mechanism for determining PD, in order to ensure that they are meeting their ethical obligations to provide accurate assessments.

Conclusion

The present analysis demonstrates that psychologists working in evaluative contexts in WC settings may be putting themselves in ethically precarious situations in their legally mandated use of the GAF to evaluate PD relating to alleged psychiatric injuries. Although Goldman et al. (1992) highlighted a primary limitation precluding the validity of the GAF, and despite the fact that others have subsequently echoed these sentiments in reference to the DSM-IV (e.g., Hilsenroth et al., 2000), the same limitation remains; specifically, a key limitation to the validity of the GAF stems from instructions for devising a single GAF score that incorporates psychiatric symptoms in addition to social and occupational functioning, with assessors instructed to take the lower (i.e., worse functioning) rating in instances in which there is a disparity among the different functioning levels. Arguably, agreement between raters may be limited by the lack of a clear definition for the construct of "functioning." As Dimsdale et al. (2010) describe, "Rather than specify all dimensions of functioning (existential, defensive/coping, sexual, occupational, etc.), the scale asks the rater for a 'gestalt' number from 1 to 100 on which to rate patients' overall functioning" (p. 515). While this suggests problems in any instance in which the GAF is utilized, this is especially true when the measure is being used in high-stakes settings, for which WC evaluations for PD clearly qualify.

The central problem in psychiatric injury evaluation in California, among other jurisdictions, has been that it is often conducted without heed to psychometric rigor and empirical justification as to why a particular mode of assessment has been chosen and whether it is appropriate for the task at hand (e.g., to rate psychiatric disability in an allegedly injured worker). Essentially, in the case of the GAF, psychologists are using a score from a measure not validated in the context for which it is being used to come to important evaluative conclusions. This unfortunate circumstance has contributed to a reputation of alleged psychiatric injuries as frivolous or downright fraudulent. As Reville et al. (2005) described, "the need for more objective and consistent ratings has been a recurring theme in California workers' compensation policy" (p. 28). Despite its prevalent use in the WC arena, there is a dearth of empirical research around the appropriateness of utilizing the GAF score in this context. Thus, there is much room for further research to specifically consider some of the reliability and validity concerns that were broached in the present paper in order to determine a more robust assessment protocol.

Although this analysis has taken a somewhat critical stance toward the current treatment of psychiatric injuries and mental health in the California WC system, the authors acknowledge that the issue of psychiatric claims in WC is a complex and controversial one. Even beyond concerns of calculated malingering or feigned exacerbation of injuries for financial gain, some researchers have argued that the relationship between financial compensation and disability is a circuitous one that must be further explored (Schatman, 2013). Although the literature continues to be divided, some claim that compensation among injured workers is associated with poorer outcomes, a phenomenon that has been conceptualized as a form of the psychoanalytic "secondary gain." Schatman explored this controversial literature in a 2013 review and concluded that the disparities in systems across states have led to "… compensation systems (that) are inconsistent, ranging from 'the good, the bad, to the ugly.'" Further, Schatman argued that attitudes toward injured workers are by and large negative such that applicants are considered malingerers until proven otherwise.

Young (2008), in discussing the challenges associated with the marriage of psychology and law in forensic settings, recommended that psychologists "stick to reasoned conclusions based on comprehensive assessments" in order to maintain the integrity of their field (p. 179). Unfortunately, the present analysis concludes that the current use of the GAF does not lend itself to reaching such reasoned conclusions—thereby making the current practice ethically, psychometrically, and clinically—if not legally—problematic.

References

45 C.F.R.§164.512(l) (2011).

Aas, I. M. (2010). Review Global Assessment of Functioning (GAF): properties and frontiers of current knowledge. *Annals of General Psychiatry, 9,* 20–31.

Allmon, A. L. (2013). Religion and the DSM: from pathology to possibilities. *Journal of Religion and Health, 52,* 538–549.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013a). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychiatric Association. (2013b). "Frequently asked questions about DSM-V implementation—for clinicians." [Web post]. Retrieved from http://www.dsm5.org/Documents/FAQ%20for%20Clinicians%208-1-13.pdf

American Psychological Association. (2010). Ethical principles of psychologists and code of conduct. Retrieved from http://apa.org/ethics/code/index.aspx

Bacon, S. F., Collins, M. J., & Plake, E. V. (2002). Does the Global Assessment of Functioning assess functioning? *Journal of Mental Health Counseling, 24,* 202–212.

Bates, L. W., Lyons, J. A., & Shaw, J. B. (2002). Effects of brief training on application of the Global Assessment of Functioning Scale. *Psychological Reports, 91,* 999–1006.

Bodlund, O., Kullgren, G., Ekselius, L., Lindström, E., & Knorring, L. (1994). Axis V—Global Assessment of Functioning Scale. *Acta Psychiatrica Scandinavica, 90,* 342–347.

Borkosky, B. G., Pellett, J. M., & Thomas, M. S. (2014). Are forensic evaluations "health care" and are they regulated by HIPAA? *Psychological Injury and Law, 7,* 1–8.

Bornstein, R. F. (2011). Toward a process-focused model of test score validity: improving psychological assessment in science and practice. *Psychological Assessment, 23,* 532–544.

Bryant v. Giani Inv. Co., 626 So.2d 390 (La. Ct. App. 1993).

Burlingame, G. M., Dunn, T. W., Chen, S., Lehman, A., Axman, R., Earnshaw, D., & Rees, F. M. (2005). Special section on the GAF: selection of outcome assessment instruments for inpatients with severe and persistent mental illness. *Psychiatric Services, 56,* 444–451.

CAL. CODE REGS. tit. 8, §43. (2009). http://www.dir.ca.gov/t8/43.html

CAL. LAB. CODE § 3208.3. (West 2011).

CAL. LAB. CODE § 4660.1(c)(1) (West 2011).

California Commission on Health and Safety and Workers' Compensation (CHSWC). (2008). Summary of system changes in California workers' compensation. Retried from http://www.dir.ca.gov/chswc/reports/chswcrptonsummarysystemchangesdraftfeb%202008.pdf

California Division of Workers' Compensation. (2005). *Schedule for rating permanent disabilities*. Sacramento, CA: Publications & Information Unit. http://www.dir.ca.gov/dwc/pdr.pdf

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290.

Clark, L. A., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment, 7,* 309–319.

Cocchiarella, L., & Anderson, G. B. J. (Eds.). (2001). *Guides to the evaluation of permanent impairment* (5th ed.). Chicago, Ill: American Medical Association.

Connell, M. A., & Koocher, G. P. (2003). HIPAA and forensic practice. *AP-LS News, 23,* 16–19.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Dep't of Corr. v. Workers' Comp. Appeals Bd., 90 Cal.Rptr.2d 716, 720 (Ct. App. 1999).

Dimsdale, J. E., Jeste, D. V., & Patterson, T. L. (2010). Beyond the global assessment of functioning: learning from Virginia Apgar. *Psychosomatics, 51,* 515–519.

Division of workers' compensation-medical unit: application for appointment as qualified medical evaluator (10, 2013), https://www.dir.ca.gov/dwc/FORMS/QMEForms/QMEForm100.pdf

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33,* 766–771.

First, M. B., & Pincus, H. A. (2002). The DSM-IV text revision: rationale and potential impact on clinical practice. *Psychiatric Services, 53,* 288–292.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7,* 248–260.

Geisinger, K. F. (2013). Reliability. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 21–42). Washington, DC: American Psychological Association.

Gold, L. H. (2014). DSM-5 and the assessment of functioning: the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0). *Journal of the American Academy of Psychiatry and the Law Online, 42,* 173–181.

Goldman, H. H., Skodol, A. E., & Lave, T. R. (1992). Revising axis V for DSM-IV: a review of measures of social functioning. *American Journal of Psychiatry, 149,* 1148–1156.

Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *The Clinical Neuropsychologist, 27,* 108–137.

Groth-Marnat, G. (2009.) Introduction. In *Handbook of psychological assessment* (5th ed., pp. 9–23.) Hoboken, NJ: Wiley.

Guyton, G. P. (1999). A brief history of workers' compensation. *The Iowa Orthopedic Journal, 19,* 106–110.

Hall, R. C. (1995). Global assessment of functioning: a modified scale. *Psychosomatics, 36,* 267–275.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 16,* 257–272.

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R., & Holdwick, D. J. (2000). Reliability and validity of DSM-IV axis V. *American Journal of Psychiatry, 157,* 1858–1863.

Hohlstein v. St. Louis Roofing Co., 49 S.W.2d 226 (Mo. Ct. App. 1932).

INDUSTRIAL MEDICAL COUNCIL, PSYCHIATRIC PROTOCOLS (1992) (amended 1993).

Kahn, M. W., Bell, S. K., Walker, J., & Delbanco, T. (2014). Let's show patients their mental health records. *JAMA, 311,* 1291–1292.

Kaufmann, P. M. (2009). Protecting raw data and psychological tests from wrongful disclosure: a primer on the law and other persuasive strategies. *The Clinical Neuropsychologist, 23,* 1130–1159.

Kennedy, J. A. (2003). *Mastering the Kennedy Axis V—a new psychiatric assessment of patient functioning*. Washington, D.C.: American Psychiatric Publishing, Inc.

Loevdahl, H., & Friis, S. (1996). Routine evaluation of mental health: reliable information or worthless 'guesstimates'? *Acta Psychiatrica Scandinavica, 93,* 125–128.

Luborsky, L. (1962). Clinicians' judgment of mental health. *Archives of General Psychiatry, 7,* 407–417.

Matsumoto A. (1994). Reforming the reform: mental stress claims under California's workers' compensation system, *27 Loy. L.A. L.* Rev. 1327–1366.

Montgomery County v. Grounds, 862 S.W.2d 35 (Tex. App. 1993).

Moos, R. H., McCoy, L., & Moos, B. S. (2000). Global assessment of functioning (GAF) ratings: determinants and role as predictors of one-year treatment outcomes. *Journal of Clinical Psychology, 56,* 449–461.

Moos, R. H., Nichol, A. C., & Moos, B. S. (2002). Global assessment of functioning ratings and the allocation and outcomes of mental health services. *Psychiatric Services, 53,* 730–737.

Narrow, W. E., & Regier, D. A. (2013). Axis V: essential supplement to the DSM-5: in reply. *Psychiatric Services, 64,* 1066–1067.

Pedersen, G., Hagtvet, K. A., & Karterud, S. (2007). Generalizability studies of the Global Assessment of Functioning—split version. *Comprehensive Psychiatry, 48,* 88–94.

Phelan, M., Wykes, T., & Goldman, H. (1996). Global function scales. In Graham Thornicroft & Michele Tansella (Eds.), *Mental health outcome measures* (pp. 15–25). Berlin Heidelberg: Springer.

Piersma, H. L., & Boes, J. L. (1997). The GAF and psychiatric outcome: a descriptive report. *Community Mental Health Journal, 33,* 35–41.

Pope, K., Butcher, J., & Seelen, J. (2006). The MMPI, MMPI–2, & MMPI–A in court: A practical guide for expert witnesses and attorneys (3rd ed.). Washington, DC: American Psychological Association.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12,* 287–297.

Reville, R. T., Seabury, S. A., Neuhauser, F. W., Burton, J. F., & Greenberg, M. D. (2005). *An evaluation of California's permanent disability rating system*. Santa Monica, CA: RAND Corporation.

Rey, J. M., Starling, J., Wever, C., Dossetor, D. R., & Plapp, J. M. (1995). Inter-rater reliability of global assessment of functioning in a clinical setting. *Journal of Child Psychology and Psychiatry, 36,* 787–792.

Riley, N. D. (2000). Mental-mental claims—placing limitations on recovery under workers' compensation for day-to-day frustrations. *Missouri Law Review, 65,* 1023.

Schatman, M. E. (2012). Workers' compensation and its potential for perpetuation of disability. In *Handbook of occupational health and wellness* (pp. 341–361). US: Springer.

Schatman, M. E., & Thoman, J. L. (2014). Cherry-picking records in independent medical examinations: strategies for intervention to mitigate a legal and ethical imbroglio. *Psychological Injury and Law, 7,* 191–196.

Schultz, I. Z. (2008). Disentangling the disability quagmire in psychological injury: part 1—disability and return to work: theories, methods, and applications. *Psychological Injury and Law, 1,* 94–102.

Schultz, I. Z., & Stewart, A. M. (2008). Disentangling the disability quagmire in psychological injury and law. *Psychological Injury and Law, 1,* 103–121.

Schwartz, G. T. (1993). Waste, fraud, and abuse in workers' compensation: the recent California experience. *Maryland Law Review, 52,* 983–1015.

Smith, G. N., Ehmann, T. S., Flynn, S. W., MacEwan, G. W., Tee, K., Kopala, L. C., & Honer, W. G. (2011). The assessment of symptom severity and functional impairment with DSM-IV Axis V. *Psychiatric Services, 62,* 411–417.

Söderberg, P., Tungström, S., & Armelius, B. Å. (2005). Special section on the GAF: reliability of Global Assessment of Functioning ratings made by clinical psychiatric staff. *Psychiatric Services, 56,* 434–438.

Sonoma State Univ. v. Workers' Comp. Appeals Bd., 48 Cal.Rptr.3d 330, 332–34 (Ct. App. 2006).

Startup, M., Jackson, M. C., & Bendix, S. (2002). The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology, 41,* 417–422.

Underwager, R., & Wakefield, H. (1993). Misuse of psychological tests in forensic settings: some horrible examples. *American Journal of Forensic Psychology, 11,* 55–75.

Üstün,T.B., Kostanjsek, N., Catterji, S., & Rehnm, J., (Ed.). (2010). *Measuring health and disability: manual for WHO disability assessment schedule WHODAS 2.0*. World Health Organization.

Vatnaland, T., Vatnaland, J., Friis, S., & Opjordsmoen, S. (2007). Are GAF scores reliable in routine clinical use? *Acta Psychiatrica Scandinavica, 115,* 326–330.

Labor and Workforce Development Agency, Department of Industrial Relations, Division of Workers' Compensation, DWC Qualified Medical Evaluator (QME) Process. (2014).

Labor and Workforce Development Agency, Department of Industrial Relations, Division of Workers' Compensation, Schedule for Rating Permanent Disabilities Under the Provisions of the Labor Code of the State of California. (2005).

Young, G. (2008). Causality and causation in law, medicine, psychiatry, and psychology: progression or regression? *Psychological Injury and Law, 1,* 161–181.