

Penile Plethysmography in Assessing Volitional Impairment in Sexually Violent Predator Evaluations

Frederick Winsmann

Harvard Medical School, Cambridge Health Alliance

[Sexual Offender Treatment, Volume 12 (2017), Issue 2]

Abstract

This article incorporates and expands upon the article "Assessing volitional impairment in sexually violent predator evaluations" published in Sexual Offender Treatment, Volume 7 (2012), Issue 1 (Winsmann, 2012). The present article describes a possible additional data source to the protocol for the assessment of volitional impairment discussed in the 2012 article by incorporating a suppression administration of the penile plethysmograph on a case-by-case basis and integrating such data into an individualized conceptualization of volitional impairment.

Keywords: civil commitment, penile plethysmograph, serious difficulty in controlling behavior, sexual arousal, sexual offender, sexually violent predator, volition, volitional impairment

This article incorporates and expands upon the article "Assessing volitional impairment in sexually violent predator evaluations" published in *Sexual Offender Treatment*, Volume 7 (2012), Issue 1 (Winsmann, 2012). This article does not replace Winsmann (2012), but, rather, this article extends Winsmann (2012). It is therefore suggested that the 2012 article be read first in order to fully understand the extension of the previous discussion in the present article.¹

The presence of a mental abnormality or personality disorder is a required element for the civil commitment, or civil management, of sexual offenders in all 21 jurisdictions in the United States that have a statute addressing such commitment or management of persons deemed by a court to be a sexually violent predator (SVP). Mental abnormalities and personality disorders are legal constructs, which may be understood in most legal schemes as mental disorders causing volitional impairment and resulting in a likelihood of sexual offense recidivism. The constitutional minimum in the United States for civil commitment under SVP statutes concerning volitional impairment, via the due process clause of the 14th Amendment of the United States Constitution, is "serious difficulty in controlling behavior" (Kansas v. Crane, 2002).² Discriminating between sexual offenders in need of civil commitment and those who do not, and apparently in order to prevent the medicalization of criminality, the U.S. Supreme Court wrote in Kansas v. Crane (2002), "[T]he severity of the mental abnormality itself, must be sufficient to distinguish the dangerous sexual offender whose serious mental illness, abnormality, or disorder subjects him to civil commitment from the dangerous but typical recidivist convicted in an ordinary criminal case" (p. 413). Accordingly, the proper examination of "serious difficulty in controlling behavior" is vitally important in order to assist in distinguishing those in need of commitment from the typical recidivist.

Volition

Volition and volitional impairment are legal constructs. Hart and Kropp (2008) stated, "[Volition] is the capacity to make choices - to form goals and then to develop, implement, evaluate, and revise plans to achieve these goals" (p. 562). Volition refers to the ability to exercise appropriate decision-making leading to reasonably healthy and legal behavioral choices (Winsmann, 2012; Winsmann, 2015). The legal elements "serious difficulty in controlling behavior" (Kansas v. Crane, 2002, p. 413), and "such an inability to control behavior" (New York Mental Hygiene Law, Chapter 27, Title B, Mental Health Act, Article 10), are expressions of the construct of *volitional impairment*, which is hereinafter referred to as VI, and VI is discussed in this article only in the context of SVP civil commitment. While volition and VI are legal constructs, the psychological concepts of impulsivity and decision-making assist in understanding these constructs.

Measuring Volitional Impairment

Impulsive behavior resulting from impairment affecting the ability to choose to engage in behavior - or to inhibit such behavior - that is not consistent with the self-interest of the individual is viewed as constitutive with regard to VI in the context of SVP civil commitment (Malamuth & Malamuth, 1999; Slavin & Kriegman, 1992; Winsmann, 2012; Winsmann, 2015; Zander, 2005). Winsmann (2012) outlined a particular protocol to assess for the presence of VI in SVP cases. The protocol, hereinafter referred to as the "VI protocol," involves an idiographic and comprehensive evaluation of the examinee and the potential use of psychological testing and neuropsychological screening measures. The protocol integrates impulsivity and decision-making, viewed as bearing theoretical relationships to the legal construct of VI, and espouses an individualized conceptualization of VI in each case. Such a conceptualization is borrowed from the literature on structured professional judgment (Douglas & Reeves, 2010). A brief review of the VI protocol and its underlying components of impulsivity and decision-making now follows.

Impulsivity is defined as an impaired ability to defer deliberation, which leads to action without sufficient cognitive and emotional mediation (Stanford et al., 2009). Impulsivity is applicable to the assessment of VI in three respects, as outlined in Winsmann (2012), and these respects include (1) the large body of literature documenting the importance of impulsivity in criminality, antisociality, psychopathy, sexual assault, and sexual impulsivity disorders (Cleckley, 1976; Farrington, Loeber, & Van Kammen, 1990; Gorenstein & Newman, 1980; Kafka, 2007; Moffitt, 1993; White et al, 1994), (2) impulsivity is a consideration in violence risk evaluations (Boer, Wilson, Gauthier, & Hart, 1997; Webster, Douglas, Eaves, & Hart, 1997), and (3) the recognition in the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013) of impulsivity as a component in the diagnosis of numerous disorders (Hucker, 1997; Melton, Petrila, Poythress, & Slobogin, 2007).

However, impulsivity alone is an insufficient explanation of VI. Impulsivity is a defining feature of a wide spectrum of psychopathology including antisociality, and it is, in short, not enough to describe VI (Vognsen & Phenix, 2004). Instead, the ability to make decisions is an additional consideration in regard to volition (Winsmann, 2012).

Decision-making refers to the ability to weigh options, consider consequences, and act in a manner consistent with, and advantageous to, the long-term interests of an individual (Bechara, Damasio, Damasio, & Anderson, 1994; Martin & Potts, 2009). Research has indicated that the inability to make decisions that are consistent with these criteria reflects an impairment in decision-making (Dias-Ferreira et al., 2009; Kim, Sohn, & Jeong, 2011). Studies support impaired decision-making involved in a range of problematic behaviors including higher motor vehicle offenses in individuals arrested for driving while intoxicated as well as links between impulsivity, in patients with borderline personality disorder, and poor outcomes on the Iowa Gambling Task, which is a measure of

decision-making (Bouchard, Brown, & Nadeau, 2012; Schuermann, Kathmann, Stiglmayr, Renneberg, & Endrass, 2011).

In considering VI in SVP cases, the question is therefore whether there currently exists for the examinee a decision-making impairment in which impulse control is problematic. The goal in assessing VI in this context, as noted above, is to create an individualized conceptualization for each examinee as part of a comprehensive evaluation. This conceptualization is borne of record reviews, interviewing, and psychological testing (Winsmann, 2012).

Accordingly, the VI protocol, as outlined in Winsmann (2012), utilizes three sources of data: (1) known behavior, both prior to and subsequent to the index offense (including from records and interviews), (2) two neuropsychological measures, and (3) self-report measures. The testing that comprises the VI protocol involves well-established neuropsychological screening measures that assess cognitive flexibility, decision-making ability, and the ability to inhibit behavior. These tests include the Wisconsin Card Sorting Test (Heaton, Chelune, Talley, Kay, & Cutiss, 1993) and the Iowa Gambling Task (Bechara, 2007). The self-report measure suggested is the Barratt Impulsiveness Scale, which is the most widely-used self-report measure of impulsiveness (Patton, Stanford, & Barratt, 1995; Stanford et al, 2009). In support of the use of such testing, research has indicated that individuals identified as having had repetitive, albeit unsuccessful, attempts to control sexual behavior performed significantly less well than a group of community controls on both the Iowa Gambling Task and the Barratt Impulsiveness Scale (Mulhauser, et al., 2014). Practitioners should rely on professional judgement and substitute, delete, or add testing as indicated.

Utilizing this data in the VI protocol is accomplished by examining test scores and considering these scores in the individualized conceptualization of VI. Determinations of VI based on testing alone should not be made. Moreover, individuals with developmental delay and/or psychosis are poor candidates for these tests, and the establishment of VI with such individuals should be conducted absent this testing unless it is established that neither clarity of thought or cognitive limitations are impediments to such administration. Finally, VI is strictly a legal construct, bearing a theoretical relationship to the behavioral concepts of impulsivity and decision-making. As such, normative data from these tests are not provided for the legal construct of VI (Winsmann, 2012).

In regard to this latter point, it is also important to emphasize that normative data from these tests are not derived from samples that are specific to sexual offenders. This is a purposeful utilization of the testing. Assessment measures, when used as part of an idiographic evaluation, are not linked to any type of offender. To attempt to do so in this context would be a mistake. The standard of behavioral control is best represented by a community sample of non-offenders. The community sample is the sample that creates the normative data for these measures, and it is the sample sought out. Using a sample of sexual offenders would be misguided: such a sample would create normative data for sexual offenders as benchmarks for this group's abilities as opposed to using community benchmarks. This thinking is not unlike the use of the "reasonable man" standard in the law, where the analysis of behavior in the eyes of the law is compared to similarly-situated individuals in the community and that which such similarity situated individuals would reasonably decide to do under similar circumstances (Winsmann, 2012).

In summarizing the utilization of this data, part of that which this author wrote in 2012 is recited here:

Ultimately, an idiographic approach is regarded as central to conducting a comprehensive evaluation. This approach does not abandon nomothetic

components, but, rather, incorporates such data into a larger picture. Structured professional judgment uses this approach in creating an individualized approach to risk when examining risk of recidivism for general violence (e.g., Historical Clinical Risk - 20) or sexual violence (e.g., Sexual Violence Risk - 20) (Douglas and Reeves, 2010). In sum, each case must be understood idiographically, and the nomothetic data that the tests produce must be integrated into such an idiographic understanding of the person. That is, an individualized conceptualization of VI is created in each case. (Winsmann, 2012, p. 8)

This VI protocol is viewed as useful because much of current evaluative practice fails to even consider volition. It was the position taken in Winsmann (2012), and again in this article, that the collection and use of scientifically-sound data concerning VI improves the data set for fact-finders to consider in SVP cases. Otherwise, fact-finders are left with very little to consider in making determinations concerning VI. The collection and proper usage of such data is viewed as a much better approach than ignoring the volitional issue altogether or relying solely on clinical judgement (Hart & Kropp, 2008). Toward achieving this goal, and improving the protocol, psychophysiological data may add clarity to the question of "serious difficulty in controlling behavior" in certain cases, and the collection and utilization of such data is therefore a possible addition to the VI protocol.

Accordingly, the VI protocol is expanded, based on research outlined in this article, by incorporating the penile plethysmograph (PPG) where indicated. The use of the PPG and measurement of penile tumescence is focused in the VI protocol on a PPG suppression administration in which examinees are exposed to potentially arousing sexual stimuli and initially instructed to respond in any manner the stimuli may elicit, and, subsequently, instructed to attempt to cognitively suppress or control sexual arousal to the same stimuli. The expansion of the VI protocol is built upon data that may be probative concerning the ability, or inability, to suppress or regulate sexual arousal. The data are deemed as possibly helpful in certain cases in creating the individualized conceptualization of VI for each examinee. This expansion aims to improve the data set in examining whether there currently exists for the examinee a decision-making impairment where impulse-control is problematic. This data is all part of determining whether evidence for the legal constructs of a mental abnormality or a personality disorder exist.

Sexual Arousal and the Link to Behavioral Control

In order to discuss such an addition to the VI protocol, a discussion of sexual arousal and the link to behavioral control first follows. Such an effort is deemed necessary before providing the practical technique of utilizing the PPG in the assessment of VI. While it is beyond the scope of this article to fully describe and discuss the literature that seeks to define sexual arousal, it is important to briefly mention several key conceptualizations and considerations concerning sexual arousal to orient the reader to the approach taken in this article to expand the VI protocol.

Sexual Arousal

Human sexual arousal has been described as a drive (Nordgren, Harreveld, van der Pligt, 2009), an emotion (Everaerd, 1989), and a motivational process (Singer & Toates, 1987). Sexual arousal conceptualized as a drive has fallen out of scholarly vogue (Pfaus, 2007), whereas sexual arousal is discussed in the contemporary psychological literature as an emotional or motivational state (Janssen, 2011).

Human behavior includes anything that a human is able to do on a mental, biological, or behavioral level (Smith, Sarason, & Sarason, 1982). Thus, the way in which sexual arousal is defined or conceptualized aside, the state of sexual arousal may be viewed, in and of itself, as a type or component of human behavior. Barlow (1977) adeptly expresses the way in which arousal is part of behavior:

"The function of behavioral assessment in an ideal world would be the direct and continuous measurement of the ... behavioral problem in the setting where the behavior presents a problem. ... In some cases the behavior cannot be conveniently produced even in contrived situations. When this happens, as in the case of sexual behavior, clinicians move back down the behavioral chain and measure sexual arousal, presumably an earlier component in the chain of sexual behavior" (as cited in Laws, 2009, p. 14).

This state of sexual arousal can be triggered by internal and external stimuli, and the existence of the state can be inferred from central (e.g., verbal), peripheral (e.g., genital), and motor responses (Janssen, 2011). Accordingly, and returning to the earlier point that it is beyond the scope of this article to fully discuss the literature concerning sexual arousal, it is nonetheless clear that sexual arousal involves behavior of the organism: a physiological change occurs in the state of sexual arousal that represents a behavior with a particular motivational goal of sexual fulfillment. This fulfillment, marked by sexual orgasmic release, represents behavior, from an evolutionary perspective, intended to propagate the species and/or achieve relational intimacy (Opperman, Braun, Clarke, and Rogers, 2014), or this behavior may alternatively be directed at paraphilic objects.³ Moreover, research supports the idea that sexual arousal can impair decision-making, which is a cognitive event, and this is important because a component of VI is impaired decision-making (Bancroft et al., 2003). As evidence buttressing this view, Laws (2009) states,

In the author's view, cognition is the motor that drives sexual arousal. It is often said that sexual arousal is a mental, not a physiological, event. From this point of view, penile erection is merely an epiphenomenon. We measure penile erection because we can, because it is an approximation of what we are seeking. We are unable to measure mental events directly, although the fMRI procedure is bringing us closer to that (p. 26).

In other words, the PPG is an analogue measurement for sexual behavior. And, sexual behavior is a sequence of events. Thus, other events in the behavioral chain must also be considered in examining sexual behavior (O'Donohue & Letourneau, 1992).

This analysis of sexual arousal is provided to justify the use of a tool to assist in examining the ability to control arousal and the data from the tool as potentially probative data in examining whether an individual can control such behavior when crafting an individualized conceptualization of VI. If evidence of the ability to control behavior exists, then it may be possible that the civil commitment of an individual with this ability to control was erroneously based on a failure to distinguish, as the Court in *Kansas v. Crane* (2002) warned, "[T]he dangerous sexual offender whose serious mental illness, abnormality, or disorder subjects him to civil commitment from the dangerous but typical recidivist convicted in an ordinary criminal case" (p. 413).

Expressed differently, civil commitment, where VI is concerned, is not appropriate if it is simply based on an individual making a bad and criminal decision to offend: civil commitment of sexual offenders in the United States, in regard to VI, is reserved for those whose current ability to control themselves is impaired. Otherwise, the medicalization of criminality occurs (Frances, 2013). This is the case because neither sexual arousal or the presence of VI are necessary for sexual offending (O'Donohue & Letourneau, 1992). The key point related to this examination of the suppression of sexual arousal, therefore, is the idea that such suppression may be a key consideration in the examination of the ability to control sexual behavior. Such an ability, or lack thereof, may be probative in examining VI and deciding upon civil commitment, which is different than criminal incarceration, which does not require a "serious difficulty in controlling behavior." However, in order to further consider suppression of arousal in the examination of the ability to control sexual behavior, further exploration of sexual arousal as a concept and the role on inhibition is indicated.

Sexual arousal and suppression

There exist numerous models outlining the process of sexual arousal in humans (Janssen, 2011), and, more recently, the idea of suppression playing a role in the human arousal process has been posited. There is evidence to support the ability to voluntarily suppress arousal (Abel, Barlow, & Blanchard, 1981; Adams, Motsinger, McAnulty, & Moore, 1992; Bancroft, 1999; Freund, 1963, 1965, 1967; Golde, Strassberg, & Turner, 2000; Henson & Rubin, 1971; Laws & Rubin, 1969; Mahoney & Strassberg, 1991; McAnulty & Adams, 1991; Quinsey & Bergersen, 1976; Quinsey & Carrigan, 1978). In fact, Bancroft and Janssen (2000) articulated a "dual control model" of sexual response, which includes suppression as one of two controls. The model posits (1) excitatory processes and (2) inhibitory processes (the dual controls) existing in the organism. The lack of arousal can be caused by either the lack of excitation or by active inhibition (i.e., suppression) of arousal (Bancroft & Janssen, 2000; Janssen and Bancroft, 2007). As Bancroft et al. (2003) stated, "The basic assumption here is that in a state of sexual arousal, normal, 'rational,' decision-making [emphasis added] is impaired; sexual arousal, and the need to experience orgasmic release, determine how the situation is handled" (p. 557). That which prevents the arousal from occurring is the suppression component of the "dual control model."⁴ Describing the evolutionary basis for the suppression control, Bancroft (1999) cited others and wrote, "Bjorklund and Kipp (1996) make the persuasive case that inhibitory mechanisms became necessary in small groups of hominids for the purposes of cooperation, group cohesion and individual political success" (p. 764). This evolutionary description of the basis for the suppression control in the "dual control model," explains, when examined from a contemporary socio-legal perspective, the suppression control as necessary to establish morally acceptable relationships, inhibit the commission of sexual offenses, and, in modernity and post-modernity at least, prevent the need for civil commitment as a sexual offender.

It is important to note that some older research utilizing the PPG does not support voluntary control over arousal (Freund, 1965, 1967). However, this inability may have been an artifact of unattraction (Henson & Rubin, 1971; Laws & Rubin, 1969), and modern PPG administrations ensure attention or data is discounted or not used altogether. This viewpoint concerning the inability to control arousal appears dated, and, in fact, sexual offenders in more modern sexual offender treatment programs often receive training to control sexual arousal as part of treatment (Laws & Holman, 1978; Laws & Marshall, 1991; Marshall, 1973; Marshall, 1979; Marshall, 2007). Such training is also undertaken in the treatment of adolescent sexual offenders (Aylwin, Reddon, & Burke, 2005). Other recent research has indicated that participants, who were pedophilic, were not able to successfully suppress sexual arousal (Babchishin, Curry, Federoff, Bradford, & Seto, 2017). However, it may be that such individuals actually did have a "serious difficulty in controlling behavior."

Research on Suppression

Administration of the PPG in suppression conditions is not novel (Mahoney & Strassberg, 1991; Winters, Christoff, & Gorzalka, 2009). Freund (1963) undertook the first empirical study of voluntary control of sexual arousal in men with results indicating a minority of men able to both generate and inhibit sexual arousal to appropriate and inappropriate stimuli. Laws and Rubin (1969) examined participants' ability to inhibit erections, and results indicated that participants had some success. These studies suffered from a methodological problem, where examinees' attention to stimuli was not ensured, while Henson and Rubin (1971) corrected for this problem and additionally found that some participants were able to exhibit a degree of suppression (Mahoney & Strassberg, 1991).

The ability to suppress sexual arousal, as measured by phallometry, differs in individuals (Abel, Barlow, & Blanchard, 1981; Babchishin et al., 2017). Emotional detachment has been found to be related to an individual's inability to suppress sexual arousal (Mahoney & Strassberg, 1991; Winters et al., 2009), and the ability to regulate emotions has been found to correlate with the ability to inhibit emotional responses to sexual stimuli (Winters et al. 2009). In addition, evidence has been found supporting the ability of instrumental conditioning to facilitate effective suppression (Rosen, 1973), and such evidence has also been found in sexual offenders with intellectual deficits (Walker, Joslyn, Vollmer, & Hall, 2014) and females (Cerny, 1978). Moreover, the inability to suppress sexual arousal has been associated with a greater likelihood of sexual risk taking in non-offending men (Babchishin et al., 2017), and sexual offenders unable to suppress arousal to child stimuli in a laboratory setting may have such difficulty outside the laboratory setting (Marshall, 2007).

There is also evidence that sexual offenders, who showed inappropriate arousal, were able to control this arousal when asked to suppress it (Avery-Clark & Laws, 1984; Freund, Watson & Rienzo, 1988; Laws & Holmen, 1978). In one particular study, sexual offender participants were able to suppress arousal to stimuli that had previously caused arousal (Hall, Proctor, & Nelson, 1988). Additionally, Abel, Barlow, and Blanchard (1981) found that most participants in a study largely comprised of sexual offenders were able to decrease arousal when asked to do so (O'Donohue & Letourneau, 1992).

Thus, taken on the whole, this research on suppression provides a platform for the view that measuring an individual's ability to suppress sexual arousal may provide important data in assessing VI (Babchishin, et al., 2017). And, this ability, or inability, to suppress arousal is not unlike the concept of self-regulation (Gross, 1998; Winters et al., 2009). Assistance in understanding this link between arousal and self-regulation is seen in the work of Janssen et al. (2016), who wrote, "Self-regulation is the capacity to control one's emotional, attentional, and/or behavioral responses. Not surprisingly, self-control is inversely related to risky decision making [emphasis added] including risky sexual behavior" (p. 2). Furthermore, in one study, cortical regions associated with emotional regulation were activated when participants were prompted to inhibit sexual arousal (Beauregard et al., 2001).

These are important points because the legal construct of VI has been linked to decision-making and impulsiveness in the Winsmann (2012) article. Thus, evidence of the ability, or inability, to suppress sexual arousal may bear directly upon the question of the ability, or inability, to control sexual behavior. And, while research noted above has shown that a large percentage of some men could not suppress sexual arousal (pedophilic offenders), it may be, as noted, that some individuals in this group, in that particular study, did indeed have a "serious difficulty in controlling behavior" (Babchishin et al., 2017), which would make these offenders arguably eligible for SVP civil commitment. A PPG suppression procedure may therefore add probative data to an individualized conceptualization of VI in certain cases and either help prevent the medicalization of criminality or

contribute evidence toward the proper civil commitment of those in need of confinement and treatment. In short, a PPG suppression procedure may serve to provide evidence concerning behavioral control, and there is research evidence to support such usage.

It is the inhibition - the suppression of arousal - measured by means of change in penile tumescence, which is a peripheral indicator of arousal, that is the target of the procedure outlined in this article. In fact, Proulx (1989) expressed the idea that penile response is the only physiological response specific to sexual arousal in men. Moreover, penile response has been described as the most sensitive index of sexual arousal and the most reliable of the physiological measures. When sexual arousal is understood as a form of behavior, evidence of suppression of sexual arousal, or lack thereof, may be evidence of control or dyscontrol of sexual behavior (Howes, 2003; Rosen & Keefe, 1978). The idea that penile erection is a totally involuntary response has faded, and penile tumescence may be understood to be under a degree of voluntary control (Winters et al., 2009). In addition, measuring control of arousal has been demonstrated in research, and, as noted above, some sexual offenders are trained to control arousal in treatment (Hatch, 1981; Laws and Holmen, 1978; Mahoney and Strassberg, 1991). Summarizing this thinking, Annon (1988) points out that one of the reasons why an individual may not reach a significant change in tumescence to a stimulus is due to the person being able to successfully control arousal (Barker & Howell, 1992).

Therefore, a proposed and possible addition to the previously published protocol for assessing VI is the administration of a suppression PPG. Such an administration may provide data regarding the ability to regulate and/or control arousal. This data, in turn, may bear on the question of whether the examinee has a "serious difficulty in controlling behavior." Accordingly, a discussion of the PPG now follows.

The PPG

Masters and Johnson (1966) offered succinct and pithy advice for those examining male sexual arousal when they suggested that the best way to make such an examination is by looking at what is happening with a man's penis. In short, stimuli causing changes in penile tumescence can be considered as representative of that which arouses the individual sexually (Barker & Howell, 1992; Laws & Holmen, 1978; Laws & Osborn, 1983). In order to make such measurements, the PPG, an instrument designed to physiologically measure sexual arousal in human males by measuring changes in penile tumescence, is utilized. It is utilized because phallometry is the best available measure of male sexual arousal (Zuckerman, 1971; Marshall, 2014). A discussion of the history, psychometric properties, administration, scoring, and interpretation of the PPG is now provided.

The summary review as that which follows is necessary and responsible because a standardized manual addressing the use of the PPG does not exist. This necessity is based on the "Responsibility of Test Users" in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The use of assessments is also addressed in the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association (hereinafter referred to as the APA Ethics Code; American Psychological Association, 2017), and this review of the PPG is also viewed as consistent with the APA Ethics Code in regard to the use of assessments. Howes (1995), in writing about the typical use of the PPG, and not an adapted technique, also provides support for this position in stating, "It is therefore incumbent upon clinicians to familiarize themselves with the relevant research in this area before embarking on any interpretation..." (p. 13).

As such, those aspects of scoring and interpretation, which are particularly germane to this analysis

of, and use of, a suppression PPG are examined in more detail as they are part-and-parcel to the suppression PPG outlined herein. That is, scoring and interpretation are examined and explained in detail because the scoring and interpretation are the bases for potentially assisting in determining whether sexual arousal exists and whether evidence of the ability to suppress sexual arousal exists. Moreover, the literature is lacking a concise summation of scoring and interpretation of the PPG, and due to the use of a suppression PPG in cases where serious liberty issues are at stake, such a summation is provided in order to allow for the proper scrutiny of the suppression protocol provided in this article. Readers are referred to the literature for additional reviews of the PPG, its history, psychometric properties, administration, scoring, and interpretation.

History of the PPG

A plethysmograph is a tool to measure variations in the volume of an organ (Laws, 2009). The PPG is utilized throughout the world, including the United States, Canada, and several European nations, to measure such change in the penis (Wilson & Miner, 2016). The instrument had its inception in the 1950s in examining sexual arousal, and, accordingly, it is the tool that has been in existence for the longest period of time for the purpose of measurement of sexual arousal in males. Relatedly, Wilson and Miner (2016) wrote, "The introduction of the phallometric method in the mid-twentieth century represented a big step forward in the understanding of male sexual arousal and preferences, particularly in regard to forensic applications" (p. 124). The PPG is built upon the "sexual preference" hypothesis (Freund, Chan, & Coulthart, 1979; Freund & Blanchard, 1981), and much of the use of the PPG has been in the forensic assessment of sexual offenders. It is in this context that the PPG is discussed in this article.

In the decades since its inception, the PPG has become, according to Murphy et al. (2015), "A standard objective measure of arousal and is considered by some researchers and clinicians to be essential in the assessment and treatment of male sex offenders and men with paraphilic interests" (p. 1). Accordingly, a proper administration of the PPG is considered a respectable scientific method, and the administration should produce reliable and valid data (Laws, 2009). In fact, Camilleri and Quinsey (2008) described the PPG as one of the best-validated tools for assessing pedophilia. In short, the PPG, when properly utilized, is viewed as a legitimate measure of human male sexual arousal (Howes, 1995; Sachs, 2000; Sachs, 2007), but there are limitations regarding its validity and reliability, which should be considered when incorporating the PPG in any psychosexual evaluation. Awareness of limitations concerning the reliability and validity of the PPG is of paramount importance, and such awareness should assist in preventing reliance on PPG data in isolation and without integrating the data into a comprehensive data set (Tong, 2007).

Psychometric Properties of the PPG

The PPG has been described as a procedure, a test, and a form of observation (O'Donohue & Letourneau, 1992). The debate about the form of assessment that the PPG represents is not entertained here. Instead, it is assumed that traditional psychometric criteria are relevant to the PPG methodology, and a brief examination of its psychometric components is undertaken. As a complete review of the psychometric properties of the PPG is beyond the scope of this article, readers are referred to the literature for other discussions of various aspects of the research in this area (e.g., Laws, 2009; Marshall & Fernandez, 2000; Marshall, Laws, & Barbaree, 1990; Merdian & Jones, 2011; O'Donohue & Letourneau, 1992; Wilson, 2016).

Research results concerning the psychometric properties of the PPG have been somewhat varied (Wilson, 2016), and standardization has been the greatest obstacle to the introduction of assessment protocols involving the PPG (Wilson & Miner, 2016). Standardization involves using the

same equipment, the same stimulus sets, the same methods of scoring, and the same methods of interpretation. While there is a great deal of similarity in the features of the different approaches to phallometry, the field has been without standardization. This lack of uniformity has created problems in psychometrically researching the instrument. Such is the case because examining data derived from different equipment, different stimuli, and different methods of interpretation creates difficulty in integrating such data (Marshall & Fernandez, 2000).

Other psychometric considerations, beyond standardization, relative to the value of the body of research concerning the use of the PPG, on the whole, include (1) the fact that a great portion of the research on the PPG utilized putatively normal males, which raises questions concerning the generalizability of this research to sexual offenders, (2) the fact that penile tumescence is but an analogue for sexual arousal, and (3) most research does not rely on individualized PPG measurements that are based on each individuals' change from flaccidity to full erection (O'Donohue & Letourneau, 1992). Similar problems are seen in other assessment measures where non-specific development samples and analogues for the behavior under study are utilized. As with other assessments, these problems have caused struggles in the field to examine the scientific soundness of the PPG, while such examination has, nonetheless, been possible and significant research has been completed.

In regard to reliability, which refers to the degree to which measurement error, instead of actual variance, has caused differences in results, there is research evidence supporting this property (O'Donohue & Letourneau, 1992). Several studies have produced adequate test-retest reliability and internal consistency (Letourneau, 2002). However, some test-retest reliability studies produced only minimal levels of reliability (Marshall & Fernandez, 2000). The reason the results did not reach higher than minimal levels appears due to the ability to suppress arousal during PPG administrations (Wilson, 2016). Studies concerning test-retest reliability have also produced substantially varying coefficients (Merdian & Jones, 2011). In regard to internal consistency, which refers to the ability of a test to produce similar results among different parts of the test and as it measures a single phenomenon, research has produced moderate to high results, while the structure of some of the studies has been criticized (Abel, Huffman, Warberg, & Holland, 1998; Frenzel & Lang, 1989; Laws, 2003; Marshall, 2006; Marshall & Fernandez, 2000; Odeschoo, 2004). Adequate internal consistency has also been found after increasing the number of stimuli in each age category and providing uniformity to stimulus voices and background (Wilson, 2016; Wison & Miner, 2016).

In regard to validity, which refers to the degree to which the instrument measures that which it is intending to measure, some research has been supportive. However, this same research has been questioned for problems in study designs (O'Donohue & Letourneau, 1992). A brief review of different types of validity concerning the PPG now follows.

Construct validity, which refers to how well the test measures the construct it was designed to measure, includes both convergent and discriminant validity. In use with child molesters, the PPG has demonstrated construct validity, while confounding considerations do not make the findings unequivocal (O'Donohue & Letourneau, 1992). Construct validity of the PPG has been studied more recently in terms of convergent validity. The PPG has done quite well when compared, for example, to the Multiphasic Sex Inventory and a self-report card sort measure of sexual interest (Laws, 2009). Quinsey, Steinman, Bergersen, & Holmes (1975) also found convergence between the PPG and a self-report measure with children and adults. In short, research consistently reports reasonable correlation coefficients between the PPG and self-report measures (Merdian & Jones, 2011). However, awareness that penile tumescence is an analogue for sexual arousal must be considered (O'Donohue & Letourneau, 1992).

In regard to discriminant validity, which refers to how well the test distinguishes groups that are expected to be different in various dimensions, research has provided stronger results in this regard with extra-familial child sexual offenders than other groups of sexual offenders (Marshall & Fernandez, 2003). Relatedly, research on the PPG has generally provided supportive evidence concerning its use in distinguishing age and gender preferences. Finally, in regard to construct validity, faking is a consideration that may limit the value of results in non-suppressed administrations. However, steps to combat faking, including semantic tracking, pneumatic seat cushions, and respiration monitors have been developed (O'Donohue & Letourneau, 1992).

Criterion validity concerns the relationship between a test and one or more outcome measures and includes predictive and concurrent validity. In regard to predictive validity, which refers to the frequency with which those who exhibit inappropriate sexual arousal actually sexually recidivate, two meta-analyses are instructive. Hanson and his colleagues found that a pedophilia index, which was calculated in phallometric testing utilized in both of these meta-analyses of the predictors of sexual reoffending, indicated that the PPG was the strongest predictor of sexual offense recidivism (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2005). Predictive validity of the PPG in regard to offenses against adults has not been as strong as that found concerning offenses against children (Marshall & Fernandez, 2000). Concurrent validity refers to a comparison between the measure in question and an outcome assessed at the same time. In regard to this type of validity with child molesters, a strong research basis exists in distinguishing these offenders from normal controls but less so in distinguishing these sexual offenders from other types of sexual offenders (Marshall, 2006; O'Donohue & Letourneau, 1992).

Concerning ecological validity, or the extent to which the data from the PPG can be generalized to real life settings, the artificial nature of the stimulus places ecological validity into question, while some research has supported such validity (Annon, 1988; Marshall, 2007). Additionally, the more the stimulus set captures the offending type of the evaluator the better the ecological validity (Merdian & Jones, 2011). Incremental validity, or the degree to which the test improves the accuracy concerning conclusions reached in relation to information from other sources, is in need of further study (Marshall, 2006; O'Donohue & Letourneau, 1992).

In regard to the specificity and sensitivity of the instrument, the PPG has done fairly well (Freund & Watson, 1991; Barsetti, Earls, Lalumiere, & Belanger, 1998; Seto, Lalumiere, & Blanchard, 2000; Wilson, 2016). Specificity, which refers to how well the test indicates appropriate arousal in an examinee who actually lacks inappropriate arousal (true negatives), appears to be sufficient with most studies reporting rates at approximately 95%. Concerning sensitivity, which refers to how well the test can detect inappropriate arousal in individuals who actually have such inappropriate arousal (true positives), the research results have varied. The range of sensitivity levels has been reported between 40% and 95% (Freund & Watson, 1991; Wilson & Miner, 2016; Wilson, 2016). Sensitivity appears stronger with child molesters than other types of offenders (Lykins et al., 2010; Marshall, 2014).

In concluding this summary of psychometric properties, it should be noted that the PPG, like any assessment methodology, has limitations. These limitations should be taken into consideration when data from a PPG assessment is interpreted and utilized. So long as such limitations are considered, the PPG can produce data that may assist in evaluating individuals if such data is utilized in conjunction with other data and integrated into a comprehensive evaluation. Despite its limitations, and as noted earlier, phallometry remains the best approach to examining sexual arousal (Marshall, 2014).

PPG Administration

The PPG measures changes in the penis by one of two primary means. The first such means of measurement is volumetric, and the second such means of measurement is circumferential. Volumetric testing involves a glass cylinder; changes in air volume are the basis for determining changes in penile tumescence. Alternatively, circumferential devices utilize strain gauges, which are comprised of silicon tubing filled with indium-gallium and fitted with electrodes that allow for analog signals to be converted to digital data in order to assess differential levels of penile tumescence (Wilson & Miner, 2016). In the current state of the science, however, the vast majority of usage and research of the PPG involves circumferential devices. Moreover, there appear to be more similarities between the two devices than differences (Howes, 1995; Marshall, 2006; Wilson & Miner, 2016).

In using the PPG, there are a number of considerations for practitioners. These considerations revolve around the different approaches that are taken in utilizing the PPG and the data it produces. The first such consideration is the choice of stimulus set. Stimulus sets are the material that may or may not stimulate sexual arousal. Stimulus sets can be audio, visual, or both audio and visual (Looman & Marshall, 2001; Marshall, 2006). In determining which stimulus set to choose, examiners should consider the variety of sexual interests of the examinee because, by way of example, sexual arousal in children will be driven by a certain set of stimuli and sexual arousal in adults will be driven by another set of stimuli (Wilson & Miner, 2016; Wilson, 2016).

Numerous stimulus sets exist for use with the PPG, and stimuli used vary along three dimensions, which include age, gender, and type of behavior. The type of behavior usually refers to level of coercion or violence (Clegg & Fremouw, 2009; Proulx, 1989). A discussion of the various stimulus sets is also beyond the scope of this article, while it is important for practitioners to justify the selection of the appropriate stimulus set and be aware of the fact that there is not one standardized stimulus set that is referred to when the PPG is discussed and examined in the literature (McPhail et al., 2017).

Scoring and Interpretation of the PPG

The next such consideration is method of scoring. The literature provides the guidance to score the PPG (Marshall & Fernandez, 2003). There are essentially four methods of scoring that are frequently utilized with the PPG. The first of these methods is the interpretation of millimeters of change in penile tumescence (raw scores), the second is percentage of full erection, the third is the creation of ipsative *z* scores, which utilize a standard deviation of one, and the fourth involves comparing responses between deviant and appropriate stimuli, which are sometimes referred to as indices (Marshall, 2006). While a full discussion of the various scoring approaches is not undertaken in this article, it is necessary to discuss scoring approaches to some extent, and how significance is determined, as these scoring approaches are particularly relevant to a suppression PPG.

A key and essential difficulty in scoring the PPG, which represents a challenge, but not an insurmountable challenge, involves low indications of arousal (Howes, 2003). This problem is borne of the fact that most examinees do not produce a full erection in response to stimuli during a PPG administration (Wilson, 2016). Another difficulty is presented by the fact that individual differences in penile characteristics can impact scoring (Furr, 1991; Wiley & Eardley, 2007). To combat both of these problems, percentage of full erection scores are often utilized (Simon & Schouten, 1991).

In order to establish that which represents the change from flaccidity to full erection, researchers have relied upon 30mm of change in penile circumference to represent a full erection (Howes, 1995;

Hunter & Goodwin, 1992; Merdian & Jones, 2007).⁵ In further examining the use of 30mm as the benchmark to represent a full erection, Howes (2003) attempted to establish normative data. This research attempted to address the issue of low arousal and the impracticality of determining individual full erection scores. By knowing the mean and standard deviation in a normal distribution, Howes was able to reach descriptive conclusions. Perhaps the most significant conclusion he reached was based in the result that 95% of full erection scores, from his study, can be expected to fall at or below 47mm. This allows for the use of a 95% confidence level in regard to PPG data without a full erection. Hence, this research was the inception of a recommendation to utilize a value of 47mm, as opposed to 30mm, as representative of full erection. Of note, Howes (2003) actually found 32.6mm as the average full erection.

Following this logic and research, the literature points to cut-scores that are described as percentage of full erection and indicative of arousal. Howes (1995) surveyed phallometric centers and found scores considered significant ranging from 5% to 30%. However, 10% of a full erection is a widely-used threshold, which is frequently utilized as the cut-score to determine the validity of a protocol (Merdian & Jones, 2011; Wilson, 2016). In fact, the literature indicates a great many practitioners and researchers utilize the 10% cut-score, including Annon (1988), Barbaree and Marshall (1989), Freund and Watson (1991), Furr (1991), Marshall, Barbaree, and Butt, (1988), Marshall and Eccles (1991), Michaud and Proulx (2009), Seto and Kuban (1996), and Quinsey and Laws (1990). Beyond the number of practitioners and researchers utilizing 10%, research by Kuban, Barbaree, & Blanchard (1999) examining 10% supports its representation as the minimum threshold of arousal (Wilson, 2016). This is the case because there exists greater agreement between volumetric and circumferential methods when this threshold is met (Coric et al., 2005). Kuban et al. (1999) summarize this point in stating, "...the dramatic increase in agreement (>.90) between methods among subjects with a response level above 10% FE [full erection] provides assurance that a critical cutoff value of 20 or 30% FE is not necessary for accuracy" (p. 356). Moreover, Lykins et al. (2010) demonstrated that a 10% of full erection cutoff creates a marked improvement in test-retest reliability. These researchers found 10% to be an acceptable minimum response to interpret circumferential phallometry.

However, 20% of full erection as the cut-score to indicate significant arousal is seen in use by researchers and practitioners. The literature outlines such utilization, and practitioners and researchers using 20% include Howes (2003), Laws and Osborn (1983), and Murphy and Barbaree (1987). Nonetheless, it is the position taken in this article that 10% of full erection is a sufficient minimum representation of arousal based on the research outlined above.

Creating percentage of full erection scores from raw scores is done based on Howes' normative data, and forty-seven millimeters as indicative of full erection. A graph, which outlines Howes' (2003) data that approximated a normal distribution, but which does not account for the moderate, positive skewness or the mild leptokurtosis in Howes' actual graph, is seen below in Figure 1. It is converted to a perfectly-normal distribution for the purposes of assisting the reader in understanding Howes' data, how it relates to the use of 47mm as indicative of full erection, and how 10% of full erection is used in this article as indicative of arousal.

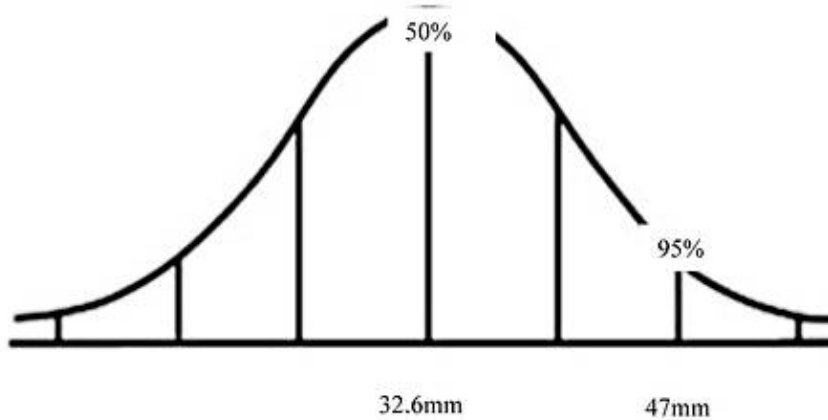


Figure 1: Frequency Distribution (n=724) of Circumferential Change Scores Converted to a Perfectly-Normal Distribution for Illustration⁶

As seen above, 95% of circumferential change scores can be expected to fall at or below 47mm. Using 10% of full erection as indicative of arousal, one can conclude that a circumferential change score greater than or equal to 4.7mm is both significant and interpretable. This conclusion is expressed in the equation that follows

$$\left(\frac{\text{Difference in arousal to a stimulus}}{47} \right) \times 100$$

(Howes, 2003)

In other words, the equation divides the difference between an examinee's baseline, non-aroused measurement of tumescence and a measurement to a stimulus in millimeters by a full erection at the 95th percentile and converts this figure into a percentage of change by multiplying by 100. I note that Marshall and Fernandez (2001) calculated the percent of full erection score by taking raw changes in the examinee's penis during trials and dividing this change by the presumed or measured full erection of the examinee and multiplying this figure by 100. So, the approach preceded Howes (2003).

The choice of using 47mm as representative of full erection is therefore supported by the use of Howes' (2003) data. The basis for this assertion is the idea that 47mm is a far more conservative number to use as the denominator in an equation of percentage of full erection. Some evaluators, as noted above, utilize 30mm. Even Howes himself found 32.6mm as the average full erection, which would cause arousal data from the PPG to be elevated above data derived from a denominator of 47. The use of 47mm may prevent the likelihood of a Type I error, where inappropriate arousal is inaccurately interpreted. This would create a false positive for data in support of a paraphilia or paraphilic disorder. The tradeoff is an increase in the possibility of a Type II error, where a true inappropriate arousal is not detected. However, the possibility of a Type II error, in this case, rather than a Type I error, is more ethically justified.

In regard to the use of z scores, which are not involved in the suppression PPG described in this article, the basis for the exclusion of this approach to scoring is due to the fact that z scores do not give a measure of the magnitude of individual arousal. In other words, separate individuals aroused

strongly to stimuli would produce similar z scores when compared to individuals aroused weakly to stimuli (Clift, Rajlic, & Gretton, 2009; Simon & Schouten, 1993). This reason is due to the fact that the data are ipsative and not normative. This problem increases the likelihood of false positives. Additionally, an index comparing appropriate and inappropriate arousal is not used in the procedure outlined in this article because the ability to suppress, rather than an examination of these types of arousal in relationship to each other, is being measured. As described below, an index is used in the suppression procedure that compares circumferential change between the unsuppressed and suppressed condition in terms of percentage of change. Hence, particular discussion of z scores and indices comparing appropriate arousal to inappropriate arousal is not made in this article.

Therefore, and in regard to the four generally prevalent approaches to scoring the PPG outlined above, this article focuses on the use of raw scores with 10% of full erection as the minimum threshold of arousal. The minimum percentage of full erection, converted to a raw score, is 4.7mm, and raw scores can be converted to percentage of full erection by means of the Howes equation.

Foundation for Updated Protocol for Assessing Volitional Impairment

As previously explained, suppression PPGs are utilized in treatment as a method to collect data in regard to the ability to control arousal. One such method is utilized at the Sand Ridge Secure Treatment Facility, which houses and treats sexual offenders in the State of Wisconsin (D. Thornton, personal communication, September 25, 2017). A suppression PPG procedure was also created by Mahoney and Strassberg (1991), which produces an "index" to establish the percentage of control in a suppressed condition when compared to the non-suppressed condition. A technique to measure responses in a suppressed condition of a PPG administration was additionally utilized in research by Babchishin et al. (2017) and Clift et al. (2009). These approaches are outlined below as they represent, in part, the foundation upon which the technique presented in this article is built.

In the approach used at the Sand Ridge Secure Treatment Facility, a normal, "baseline" PPG is administered followed by a suppression PPG if the baseline PPG indicated a stronger response to an offense-related stimulus segment compared to the examinee's strongest response to a normative segment. In the suppression condition, the instructions are, in essence, for the examinee to attend to the stimuli and use mental strategies to control any arousal response to offense-related stimuli. Data are analyzed in the context of treatment and in terms of whether the person could learn to suppress their arousal to offense-related stimuli. This ability to suppress is regarded as achieved once responses to the offense-related stimuli were all less than 5mm. (D. Thornton, personal communication, September 25, 2017).⁷

The procedure utilized by Mahoney and Strassberg (1991) involved viewing a total of five vignettes. These vignettes were used in a baseline session, which is called the "sexual/honest" condition, where examinees were instructed to "respond as you normally would" (Mahoney & Strassberg, 1991, p. 6). If at least "moderate" increases in penile circumference were achieved in the baseline session, a suppression condition was attempted with the same five vignettes. In this suppression condition, called the "sexual fake" condition, examinees were instructed to, according to Mahoney and Strassberg (1991), "Respond as though the segment were not arousing at all...suppress any arousal you may feel" (p. 6).

Data were analyzed by calculating three ratios, which included a suppression index, a development index, and a fake index. The suppression index is conceptualized as an estimate of suppressed arousal relative to baseline arousal in the "sexual/honest" condition. Creation of the index is

accomplished by dividing an examinee's highest response in the "sexual/fake" condition by the examinee's highest response in the "sexual/honest" condition and subtracting this figure from one. The authors discuss a suppression index of 0.30 as an example, which indicates a 30% reduction in arousal. The development index and the fake index are less relevant to the focus of this article and are not elaborated upon (Mahoney & Strassberg, 1991).

In the research conducted by Babchishin et al. (2017), two PPG sessions were completed, and indices comparing appropriate to inappropriate arousal were utilized. The first constituted the "normal" condition, and the second constituted the "suppression" condition. The suppression instructions were, "For the next series of tapes, continue to listen to them as you have been doing; but, from now on, try to suppress your sexual arousal. Use whatever mental means you wish to prevent yourself from feeling any sexual excitement" (p. 675). Data, including indices, were analyzed using the "clinically significant change approach" (Jacobsen, Follette, & Revenstorf, 1984), which assesses the reliability of the change between the normal and suppression conditions and classified participants into four categories.

Clift et al. (2009) examined the ability to control arousal in sexual offenders including the relationship between such control and sexual offense recidivism. This particular study examined this ability in adolescent sexual offenders and results indicated, among other findings, a positive correlation between an inability to control arousal and increased sexual offense recidivism. The approach utilized compared indices that were created from the PPG data. Such indices were crafted from examining penile circumference, measured in millimeters of change, and converting the change measured in millimeters to z scores where such scores represent arousal generated by normophilic stimuli relative to paraphilic stimuli. Such indices were compared between aroused and suppressed conditions. There do not appear to have been specific instructions given to examinees regarding how to suppress arousal in this study.

In regard to the above procedures, both Sand Ridge, and Babchishin et al. (2017) utilized polygraph data as well as other measures. Mahoney and Strassberg (1991) utilized two questionnaires, the Betts QMI Scale and the Gordon Test of Visual Imagery Control, in order to assess mental imagery abilities and to facilitate detumescence between trials (Richardson, 1969). I also note that Clift et al. (2009) focused on risk of sexual offense recidivism. In addition, Babchishin et al. (2017), Clift et al. (2009), and Mahoney and Strassberg (1991) utilized these procedures in the context of research, while Sand Ridge utilizes these procedures in the context of treatment.

Expansion of VI Protocol

The PPG utilized and discussed is this expansion of the protocol is circumferential rather than volumetric.⁸ Significance in arousal is determined by a minimum of 4.7mm of stretch in the penile strain gauge beyond the examinee's baseline.

During a typical PPG administration, the examinee is presented with various sexual stimuli consisting of different ages and genders. The stimulus set consists of "trials," which are individual exposure periods to stimuli.⁹ The examinee's baseline is established at the outset of the PPG administration and is re-established at the beginning of each trial. The content of the stimuli is visual, audio, or both visual and audio. Countermeasures are utilized in order to obtain data regarding possible confounds or errors related to detected sexual arousal. Examinees are also instructed to indicate when certain inappropriate elements are present in trials in order to ensure attention.

In this expansion of the VI protocol, the first step is to choose an appropriate PPG stimulus set. The

stimulus set for any PPG is chosen, in part, based on the presentation and history of the examinee. The Real Child Voices Stimulus Set, also known as the Burke-Musolf stimulus set, and the Marshall Stimulus Set are common examples (Murphy et al., 2015). Each stimulus set has numerous trials with stimuli that focus on different possible arousal interest.

The update to the VI protocol utilizes a selection of stimulus set trials from these or other stimulus sets used with the PPG. This update involves a baseline PPG administration and a suppression PPG administration of these select trials. This addition to the VI protocol is, hereinafter, referred to as the "VI-PPG."

A total of three PPG trials constitute the baseline VI-PPG administration, and a complete baseline, unsuppressed administration is completed first, which is followed by the suppressed administration. If, for example, the examinee exhibited a history of child offenses, three inappropriate trials are chosen that involve children. If a significant degree of unsuppressed arousal was obtained (greater than or equal to 4.7mm of stretch) for particular trials, these and only these particular trials are re-administered in the suppression condition. If significant arousal was not obtained for at least two trials in the baseline administration, the suppression condition is not administered and data collection for the VI-PPG is abandoned. It is abandoned because absent any significant arousal measurement the control of such arousal cannot be undertaken, and absent the ability to replicate the measurement of this arousal by means of at least two trials reaching baseline arousal significance the protocol is deemed to have insufficient data.¹⁰

Therefore, for a suppression protocol to be effective, and to create a valid set of data in regard to suppression, the examinee must be cooperative because absent arousal data the VI-PPG will be inconclusive (Abel, Barlow, Blanchard, & Guild, 1977). Nonetheless, the lack of arousal from this set of PPG trials may be utilized as part of a comprehensive evaluation in order to assist in ruling out arousal patterns and possible diagnostic categories. However, utilization of PPG data for the purposes of determining the presence of current arousal in order to assist in diagnosis should not be made unless the evaluator has sufficient data to do so. In addition, should examinees display historical arousal to more than one inappropriate sexual object (e.g., offenses against children and adults), two separate VI-PPG administrations should be conducted comparing arousal and control to each stimulus type.

Typical instructions are given to the examinee for the baseline administration. In regard to the suppression administration, examinees are instructed to use any mental techniques that may have been learned in treatment and any other mental techniques to control and suppress arousal. Examinees are additionally instructed not to attempt to suppress by means of physical interventions, such as clenching muscles or holding breath, because such techniques are not cognitive in nature, and because such techniques interfere with the PPG measurements. The control of sexual arousal that is of utility in examining the ability to control sexual behavior in the community involves mental strategies that will assist in decision-making and help control impulses.

If at least two trials in the baseline administration reached at least 4.7mm of stretch, and to analyze the data for change and potential evidence of the ability to suppress inappropriate arousal, a suppression index is created. The suppression index is created by analyzing the difference between baseline and suppressed conditions including any increase in arousal in the suppressed condition. The suppression index measures the reduction in arousal and controls for variations in penis size as each individual's percentage of change is calculated.

The VI-PPG is based on the work of others, particularly Mahoney & Strassberg (1991), who utilized an index that served to calculate the percentage of change between normal and suppressed

conditions. One advantage to the VI-PPG is found in the data analysis: the overall suppression index is an average of the trials instead of comparing peak responses from separate individual trials, which was the procedure used by Mahoney and Strassberg (1991). Another advantage over the approaches of others is the use of inappropriate stimuli in the VI-PPG. Mahoney and Strassberg (1991) do not appear to have used inappropriate stimuli, and Babchishin et al. (2017) and Clift et al. (2009) utilized inappropriate stimuli but analyzed suppression by means of an index that involved normophilic and paraphilic arousal. Finally, the VI-PPG relies on the normative data discussed above establishing 4.7mm as the minimal level of increased tumescence demonstrative of arousal. Babchishin et al. (2017) do not mention the tumescence threshold relied upon as indicative of arousal other than utilizing less than 2.5mm of stretch as an exclusionary criterion for their study. The procedures by Clift et al. (2009) and Mahoney and Strassberg (1991) do not appear to have used any tumescence threshold as indicative of arousal.

In summary, an attempt was made to build a VI-PPG based on extant research and make improvements by (1) examining data averages, (2) relying on the normative figure of 4.7mm as the minimum threshold of arousal, (3) utilizing inappropriate stimuli, and (4) avoiding the use of indices comparing normophilic to paraphilic data, which is seen to confound rather than clarify the suppression question. In regard to this fourth improvement, appropriate arousal is not part of the VI-PPG index for this reason, but the presence of appropriate arousal, and whether the examinee displays a stronger arousal response to inappropriate stimuli, may be relevant in crafting an individualized conceptualization of VI in certain cases. In such cases, practitioners should obtain sufficient PPG data or rely on other data that are relevant to appropriate arousal.

As an example of applying the VI-PPG as described in the narrative above, consider the hypothetical case of an individual convicted of a contact sexual offense against a female prepubescent child, who is facing civil commitment, and who carries the diagnosis of Pedophilic Disorder, Exclusive type, Sexually attracted to females. In this case, the VI-PPG is utilized to potentially assist in clarifying whether there exists a "serious difficulty in controlling behavior" borne of a nexus between Pedophilic Disorder and such a serious difficulty. Accordingly, three trials constituting arousal to prepubescent females are utilized in this example. Table 1 below is completed with fictitious data for the purposes of readers' visualizing and further understanding how the VI-PPG would be utilized.

Table 1: Sample VI-PPG Data

	Baseline PPG	Suppression PPG	Suppression
	Peak Unsuppressed Arousal	Peak Suppressed Arousal	Index
Trial 1 – Child Female 1	<i>10.41mm</i>	4.51mm	0.57
Trial 2 – Child Female 2	<i>8.88mm</i>	5.03mm	0.43
Trial 3 – Child Female 3	<i>8.90mm</i>	6.10mm	0.31
Mean	<i>9.40mm</i>	5.21mm	0.44

In regard to the data above, all unsuppressed trials reached significance of at least 4.7mm of stretch, so the data were analyzable. During the baseline administration, the examinee achieved results of 10.41mm, 8.88mm, and 8.90mm for trials one, two, and three, respectively. The mean of these data was 9.40mm. During the suppression condition, the examinee achieved results of

4.51mm, 5.03mm, and 6.10mm for trials one, two, and three, respectively. There was a reduction in the mean from 9.40mm to 5.21mm. The individual suppression index for each trial is calculated by the following equation

$$1 - (\text{peak suppressed arousal} / \text{peak unsuppressed arousal})$$

The individual trial suppression indices are therefore crafted as follows

$$\text{Trial 1} = 1 - (4.51 / 10.41) = 0.57$$

$$\text{Trial 2} = 1 - (5.03 / 8.88) = 0.43$$

$$\text{Trial 3} = 1 - (6.10 / 8.90) = 0.31$$

The overall suppression index is the mean of the three individual suppression indices and is calculated by the following equation¹¹

$$(\text{suppression index trial 1} + \text{suppression index trial 2} + \text{suppression index trial 3}) / 3$$

The overall suppression index from the data in Table 1 is therefore crafted as follows

$$(0.57 + 0.43 + 0.31) / 3 = 0.44$$

This index represents a 44% reduction in arousal on average.

In utilizing the suppression index, the mean reduction in arousal of 44% should be evaluated qualitatively and integrated into an individualized conceptualization of VI, while an individualized conceptualization of VI should not be made by means of the VI-PPG alone. In order to assist in integrating and make meaning of this index, it is suggested that examiners obtain and consider three pieces of data in particular concerning the VI-PPG. These data are obtained subsequent to the VI-PPG administration and include the examinee's account of (1) the cognitive strategies or techniques used to attempt to control arousal, (2) how such arousal-control strategies may impact impulse control and decision-making regarding behavioral actions, and (3) how such strategies would be employed in real life situations. Additionally, if there was an increase in arousal in any suppressed conditions, the examinee should be questioned about this outcome.

Evaluators should rely upon professional judgement in determining how to integrate the index and this information into an individualized conceptualization of VI. As noted earlier, such usage is not dissimilar to how risk factors and other data are integrated in a structured professional judgment risk evaluation (Douglas & Reeves, 2010). And, if more than one VI-PPG was conducted, due to a history of multiple objects of sexual arousal, this data must also be integrated and considered. The VI-PPG index is, therefore, utilized qualitatively as a component of the VI protocol, and the VI protocol, which crafts an individualized conceptualization of VI, is but a component of an idiographic

understanding of the person and a comprehensive evaluation.

Accordingly, the VI-PPG data may be probative in considering "serious difficulty in controlling behavior." In sum, it involves an analysis, as part of the individualized conceptualization of VI, of how arousal, and the ability or inability to cognitively suppress arousal, plays a role in the examinee's decision-making process and ability to control sexual impulses. Moreover, as noted, the VI-PPG serves as practical evidence of control of sexual behavior, or the lack thereof, while there are limitations in equating evaluative data with community behavior.

The VI-PPG data may be particularly useful in cases where there is conflicting data concerning "serious difficulty in controlling behavior," and in cases where that data does not allow an evaluator to reach a conclusion to a reasonable degree of certainty. The VI-PPG may also be of particular assistance, in certain cases, where the higher standard of "such an inability to control behavior" is at issue as is the case in the State of New York. These points notwithstanding, professional judgement, and case-by-case determination, must be utilized and undertaken in deciding upon the use of the VI-PPG.

Limitations and Strengths

There are limitations inherent in any attempt to make behavioral measurement including the VI-PPG addition to the VI protocol. In this regard, there are the general limitations of the PPG instrument itself; the reader is referred to the literature for further outlines of such limitations (e.g., Marshall, 2006; Murphy & Barbaree, 1994; O'Donohue & Letourneau, 1992; Wilson, 2016). Nonetheless, perhaps the most significant limitation in this regard is the lack of standardization of phallometric testing in regard to stimulus sets and scoring (Marshall & Fernandez, 2000). Inherent in this first limitation concerning the PPG generally is the idea that some sexual offenses may be driven by nonsexual motivation (Marshall & Fernandez, 2003; Meridian & Jones, 2011; O'Donohue & Letourneau, 1992).

A second such limitation is the inability to find any utility to the data should the initial unsuppressed administration not provide at least 4.7mm of stretch from baseline in at least two trials. Essentially, this causes the VI-PPG administration to be data that one cannot utilize concerning "serious difficulty in controlling behavior," and the time spent attempting the VI-PPG would be lost. A third limitation is the brevity of the VI-PPG. Limiting the administration to only three trials limits the data collected, and, hence, may limit the ability to glean the fullest picture of evidence concerning arousal control. However, this limitation may be evaluated on a case-by-case basis, and trials can be added if necessary. A fourth limitation is the generalizability of evaluative data from the VI-PPG to real-life situations in the community (Barker & Howell, 1992; Meridan & Jones, 2011). This point notwithstanding, some research supports the correlation between laboratory data and community-based behavior and the PPG (Annon, 1988; Marshall, 2007). Finally, I note that another factor that may limit the utility of the results from the VI-PPG is the possibility of habituation. That is, examinees will be exposed to the very same stimuli with a relatively short period of time between such exposures. Having already been exposed to the stimuli presented in the suppression condition, the effect of habituation may potentially confound results.

One strength of the VI-PPG concerns its practical ability to possibly provide data that bears upon the question concerning the ability to control behavior. There is some evidence to support that these data are analogous for non-laboratory arousal, as noted above, and, therefore, creates a demonstration of that which examinees will face in the community. The ability, or inability, to make decisions and control impulses lies at the heart of the VI protocol, and, accordingly, if an examinee can demonstrate the ability to control arousal by means of cognitive suppression, the presence of a

"serious difficulty in controlling behavior" can be better assessed by incorporating this scientific evidence.

Another strength concerns the VI-PPG assisting in making determinations about SVP civil commitment based on current functioning. That is, current data from the VI-PPG can be utilized in opining about SVP civil commitment criteria. Making decisions about civil commitment based largely on historical data leads to serious concerns about the sufficiency of opinions reached (Weinberfer, Sreenivasan, Azizian, & Garrick, 2018).

Caveats and Cautions

It is important that the information that may be gleaned from the VI-PPG be appropriately utilized. Accordingly, it is necessary to point out caveats and cautions to the utilization of the VI-PPG.

First, finding evidence of inappropriate sexual arousal and/or an inability to suppress such inappropriate arousal does not, taken in isolation and absent other data from a comprehensive evaluation, equate to "serious difficulty in controlling behavior" or the likelihood of sexual offense recidivism. Alternatively, the lack of evidence of inappropriate sexual arousal and/or the ability to suppress such inappropriate arousal does not, taken in isolation and absent other data from a comprehensive evaluation, equate to the lack of "serious difficulty in controlling behavior" or the lack of likelihood of sexual offense recidivism. Formulating how each individual makes decisions, and inhibits impulses, are the key components to an individualized conceptualization of VI, and such a conceptualization is not made based on VI-PPG data alone, or, for that matter, any testing data alone. Furthermore, risk of sexual offense recidivism is a separate component of an SVP evaluation.

Second, it is the change from the normal to the suppression condition that may be probative in regard to VI. Therefore, even if an examinee continues to show at least 4.7mm of circumferential change in the suppression condition, this should not be equated with "serious difficulty in controlling behavior" or a likelihood of sexual offense recidivism.

I also note that the PPG should never be used as a stand-alone measure to make a diagnosis of a paraphilic disorder or determinations of risk to reoffend sexually, and data from the PPG should never be treated as proof of past, current, or future behavior. PPG data assists in reaching conclusions in regard to sexual arousal when considered in concert with other data utilized (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Simon & Schouten, 1991).

In line with these caveats and cautions, Laws (2009) outlines inappropriate applications of the PPG, and they are recounted here, at the close of this article, to remind readers that the assessment procedure must be used with care. These inappropriate applications include (1) the use of erection responses to determine whether someone has committed a specific sexual offense, (2) the use of erection responses as the sole criterion to determine release from custody or treatment, (3) the use of erection response to screen populations for potential sexual offenders, and (4) the use of erection response as the sole source of data and absent other data in evaluations.

In short, by way of summary and re-emphasis, the VI-PPG is seen as a potentially-helpful component of the VI protocol, and the VI protocol is seen as a potentially-helpful protocol, which may be probative, in examining the question of "serious difficulty in controlling behavior." Thus, data taken in isolation from the VI protocol should not be equated with a presence or lack of a "serious difficulty in controlling behavior."

Future Considerations

Concerning future considerations vis-à-vis the VI-PPG, exploration of the varying effects of different stimulus modalities (audio, visual, or both), instructional sets (instructions given to the examinee in normal and suppressed conditions), and the stimulus sets chosen may be helpful (Abel, Blanchard, and Barlow, 1981). In addition, teaching suppression strategies to examinees may affect outcomes on the VI-PPG, and examining different types of strategies and the differences between those taught to suppress and those not taught to suppress may provide helpful information (Dougher, 1995).

In concluding this article, it is important to point-out that the PPG, when used properly, is a valuable, scientifically-sound measurement instrument. However, it must be used carefully - particularly the context of civil commitment, in which the public protection interests of the State and the liberty interests of the individual hang in the balance.

References

1. Abel, G. G., Barlow, D. H., Blanchard, E. B., Guild, D. (1977). The components of rapists' sexual arousal. *Archives of General Psychiatry*, 34, 895-903.
2. Abel, G. G., Barlow, D. H., & Blanchard, E. B. (1981). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set and stimulus content on the objective. *Behavior Research and Therapy*, 19, 25-33.
3. Abel, G. G., Huffman, J., Warberg, B., & Holland, C. (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: A Journal of Research and Treatment*, 10(2), 81-95.
4. Adams, H. E., Motsinger, P., McAnulty, R. D., & Moore, A. L. (1992). Voluntary control of penile tumescence among homosexual and heterosexual subjects. *Archives of Sexual Behavior*, 21, 17-31.
5. American Educational Association, American Psychological Association, & the National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
6. American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders: Fifth Edition*. Washington, D.C.: Author.
7. American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct (2002, Amended June 1, 2010 and January 1, 2017)*. Retrieved from <http://www.apa.org/ethics/code/index.aspx>
8. Aylwin, A. S., Reddon, J. R., & Burke, A. R. (2005). Sexual fantasies of adolescent male sex offenders in residential treatment: A descriptive study. *Archives of Sexual Behavior*, 34, 231-239.
9. Annon, J. S. (1988). Reliability and validity of penile plethysmograph in rape and child molestation cases. *American Journal of Forensic Psychology*, 6(2), 11-26.
10. Avery-Clark, C. & Laws, D. R. (1984). Differential erection response patterns of child sexual abusers to stimuli describing activities with children. *Behavior Therapy*, 15, 71-83.
11. Babchishin, K. M., Curry, S. D., Federoff, J. P., Bradford, J., & Seto, M. C., 2017. Inhibiting sexual arousal to children: Correlates and its influence on the validity of penile plethysmography. *Archives of Sexual Behavior*, 46, 671-684.
12. Bancroft, J. (1999). Central inhibition of sexual response in the male: A theoretical perspective. *Neuroscience and Biobehavioral Review*, 23, 763-784.
13. Bancroft, J. (2000). Individual differences in sexual risk taking by men: A psychosociobiological approach. In J. Bancroft (Ed.), *The role of theory in sex research* (pp.

- 177-212). Bloomington, IN: Indiana University Press.
14. Bancroft, J. & Janssen, E. (2000). The dual control model of male sexual response: a theoretical approach to centrally mediated erectile dysfunction. *Neuroscience and Biobehavioral Reviews*, 24, 571-279.
 15. Bancroft, J., Janssen, E., Strong, D., Carnes, L., Vukadinovic, Z., & Long, J. S. (2003). Sexual risk-taking in gay men: the relevance of sexual arousability, mood, and sensation seeking. *Archives of Sexual Behavior*, 32(6), 555-572.
 16. Barbaree, H. E. & Mewhort, D. J. K. (1993). The effects of the z score transformation of measures of relative erectile response strength: A re-appraisal. *Behavior Research Therapy*, 32(5), 547-558.
 17. Barbaree, H. E. & Marshall, W. L. (1989). Erectile responses among heterosexual child molesters, father-daughter incest offenders, and matched non-offenders: Five distinct age-preference profiles. *Canadian Journal of Behavioral Science*, 21, 70-82.
 18. Barbaree, H. E., Marshall, W. L., & Lanthier, R. D. (1979). Deviant sexual arousal in rapists. *Behavior Research and Therapy*, 17, 215-222.
 19. Barker, J. G. & Howell, R. J. (1992). The plethysmograph: A review of recent literature. *The Bulletin of the American Academy of Psychiatry and the Law*, 20(1), 13-25.
 20. Barlow, D. H. (1977). Assessment of sexual behavior. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (pp. 461-508). New York, NY: John Wiley & Sons, Ltd.
 21. Barsetti, I, Earls, C. M., Lalumière, M. L., Bélanger, N. (1998). The differentiation of intrafamilial and extrafamilial heterosexual child molesters. *Journal of Interpersonal Violence*, 13(2), 275-286.
 22. Bechara, A. (2007). *Iowa Gambling Task (IGT)*. Lutz, FL: Psychological Assessment Resources.
 23. Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50, 7-15.
 24. Boer, D. P., Wilson, R. J., Gauthier, C. M., & Hart, S. D. (1997). Assessing risk of sexual violence: Guidelines for clinical practice. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 326-342). New York, NY: The Guilford Press.
 25. Bouchard, S. M, Brown, T. G., & Nadeau, L. (2012). Decision-making capacities and affective reward anticipation in DWI recidivists compared to non-offenders: A preliminary study. *Accident Analysis and Prevention*, 45, 580-587.
 26. Camilleri, J. A. and Quinsey, V. L. (2008). Pedophilia. In R. D. Laws & W. T. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (pp. 183-212). New York, NY: The Guilford Press.
 27. Cerny, J. A. (1978). Biofeedback and the voluntary control of sexual arousal in women. *Behavior Therapy*, 9, 847-855.
 28. Clegg, C. & Fremouw, W. (2009). Phallometric assessment of rapists: A critical review of the research. *Aggression and Violent Behavior*, 14, 115-125.
 29. Cleckley, H. C. (1976). *The mask of sanity* (5th ed.). St Louis, MO: Mosby.
 30. Clift, R. J., Rajlic, G., & Gretton, H. M. (2009). Discriminative and predictive validity of the penile plethysmograph in adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 21(3), 335-362.
 31. Coric, V., Feuerstein, S., Fortunati, F., Southwick, S., Temporini, H., & Morgan, C. A. (2005). Assessing sex offenders. *Psychiatry (Edgemont)*, 2(11), 26-29.
 32. Dias-Ferreira, E., Sousa, J. C., Melo, I., Morgado, P., Mesquita, A. R., Cerqueira, J. J., Costa, R. M., & Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. *Science*, 325, 621-625.

33. Dougher, M. J. (1995). Behavioral techniques to alter sexual arousal. In B. K. Schwartz & H. R. Cellini (Eds.), *The sex offender: Corrections, treatment and legal practice* (pp. 1-15). Kingston, NJ: Civic Research Institute.
34. Douglas, K. S. & Reeves, K. (2010). HCR-20 violence risk assessment scheme: Rationale, application and empirical overview. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 147-185). New York, NY: Taylor & Francis Group.
35. Everaerd, W., (1989). Commentary on sex research: sex as an emotion. *Journal of Psychology and Human Sexuality*, 2, 3-15.
36. Farrington, D. P., Loeber, R., & Van Kammen, W. (1990). Long-term criminal outcomes of hyperactivity-impulsivity-attention deficit and conduct problems in childhood. In L. N. Robins & M. Rutter (Eds.), *Straight and devious pathways from childhood to adulthood* (pp. 62-81). Cambridge, England: Cambridge University Press.
37. First, M. B. & Halon, R. L. (2008). Use of DSM paraphilia diagnoses in sexually violent predator commitment cases. *Journal of the American Academy of Psychiatry and the Law*, 36, 443-454.
38. Frances, A. (2013). *Saving normal: An insider's revolt against out-of-control psychiatric diagnosis, DSM-5, big pharma, and the medicalization of ordinary life*. New York, NY: HarperCollins Publishers.
39. Frenzel, R. R. & Lang, R. A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child abusers. *Annals of Sex Research*, 2, 255-275.
40. Freund, K. (1963). A laboratory method for diagnosing homo- and hetero-erotic interest in the male. *Behavior Research and Therapy*, 1, 85-93.
41. Freund, K. (1965). Diagnosing heterosexual pedophilia by means of a test for sexual interest. *Behavior Research and Therapy*, 3, 229-234.
42. Freund, K. (1967). Diagnosing homo- and heterosexuality and erotic age preference by means of psychophysiological test. *Behavior Research and Therapy*, 5, 209-228.
43. Freund, K & Blanchard, R. (1981). Assessment of sexual dysfunction and deviation. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (2nd ed.) (pp. 427-455). New York, NY: Pergamon Press.
44. Freund, K. & Watson, R. (1991). Assessment of the sensitivity and specificity of a phallometric test: An update of phallometric diagnosis of pedophilia. *Psychological Assessment*, 3, 254-260.
45. Freund, K., Chan, S., & Coulthart, R. (1979). Phallometric diagnosis with "nonadmitters." *Behavior Research and Therapy*, 17, 451-457.
46. Freund, K., Watson, R., & Rienzo, D. (1988). Signs of feigning in the phallometric test. *Behavior Research and Therapy*, 26(2), 105-112.
47. Furr, K. D. (1991). Penis size and magnitude of erectile change as spurious factors in estimating sexual arousal. *Annals of Sex Research*, 4(3-4), 265-279.
48. Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: A new perspective and a model for research. *Psychological Review*, 87, 301-315.
49. Golde, J. A., Strassberg, D. S., & Turner, C. M. (2000). Psychophysiological assessment of erectile response and its suppression as a function of stimulus media and previous experience with plethysmography. *The Journal of Sex Research*, 37(1), 53-59.
50. Hall, G. C. N., Proctor, W. C., & Nelson, G. M. (1988). Validity of physiological measures of pedophilic sexual arousal in a sexual offender population. *Journal of Consulting in Clinical Psychology*, 56, 118-122.
51. Hanson, R. K. & Bussiere, M. T. (1998). Predicting relapse: a meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.
52. Hanson, R. K. & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 72, 1154-1163.

53. Hart, S. D., & Kropp, P. R. (2008). Sexual deviance and the law. In D. R. Laws & W. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (2nd Ed.) (pp. 37-60). New York, NY: Guilford.
54. Heaton, R. K, Chelune, G. J., Talley, J. L., Kay, G. G., & Cutiss, G. (1993). *Wisconsin Card Sorting Test Manual, Revised and Expanded (WCST)*. Lutz, FL: Psychological Assessment Resources.
55. Henson, D. E. & Rubin, H. B. (1971). Voluntary control of eroticism. *Journal of Applied Behavior Analysis*, 4, 37-44.
56. Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment*, 7(1), 9-24.
57. Howes, R. J. (2003). Circumferential change scores in phallometric assessment: Normative data. *Sexual Abuse: A Journal of Research and Treatment*, 15(4), 365-375.
58. Hucker, S. J. (1997). Impulsivity in DSM-IV impulse-control disorders. In C. D. Webster and M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 195-211). New York, NY: The Guilford Press.
59. Hunter, Jr., J. A. & Goodwin, D. W. (1992). The clinical utility of satiation therapy with juvenile sex offenders: Variations and efficacy. *Annals of Sex Research*, 5, 71-80.
60. Jacobsen, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.
61. Janssen, E., & Bancroft, J. (2007). The dual control model: The role of sexual inhibition and excitation in sexual arousal and behavior. In E. Janssen (Ed.), *The psychophysiology of sex* (pp. 197-222). Bloomington, IN: Indiana University Press.
62. Janssen, E. (2011). Sexual arousal in men: A review and conceptual analysis. *Hormones and Behavior*, 59, 708-716.
63. Janssen, E., Hoffmann, H. L., Goodrich, D., & Wilson, M. R. (2016). The effects of alcohol on self-regulation of sexual arousal in sexually compulsive men who have sex with men. *Sex Addict Compulsivity*, 23(4), 313-323.
64. Joslyn, P. R. & Vollmer, T. R. (2014). Differential suppression of arousal by sex offenders with intellectual disabilities. *Journal of Applied Behavior Analysis*, 47, 639-644.
65. Kafka, M. P. (2007). Sexual impulsivity disorders: Psychiatric "orphans." *Psychiatric Times*, 24(14), 15-17.
66. *Kansas v. Crane*, 534 U.S. 407, 413 (2002)
67. *Kansas v. Hendricks*, 521 U.S. 346 (1997)
68. Kim, Y. T., Sohn, H., & Jeong, J. (2011). Delayed transition from ambiguous to risky decision making in alcohol dependence during Iowa Gambling Task. *Psychiatry Research*, 190, 297-303.
69. Kuban, M., Barbaree, H. E., & Blanchard, E. (1999). A comparison of volume and circumference phallometry: Response magnitude and method agreement. *Archives of Sexual Behavior*, 28, 345-359.
70. Laws, D.R. (2003). Penile plethysmography: Will we ever get it right? In T. Ward, D. R. Laws, & S. M. Hudson (Eds.), *Sexual deviance: issues and controversies* (pp. 82-102). Thousand Oaks, CA: SAGE Publications Ltd.
71. Laws, D. R. (2009). Penile plethysmography: Strengths, limitations, innovations. In D. Thornton & D. R. Laws, D. R. (Eds.) *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp.7-29). Chichester, West Sussex: John Wiley & Sons.
72. Laws, D. R., & Holmen, M. L. (1978). Sexual response faking by pedophiles. *Criminal justice and Behavior*, 5, 343-356.
73. Laws, D. R. & Rubin, H. B. (1969). Instruction control of an autonomic sexual response. *Journal of Applied Behavior Analysis*, 2, 93-99.

74. Laws, D. R. & Marshall, W. L. (1991). Masturbatory reconditioning with sexual deviates: An evaluative review. *Advances in Behavior and Research Therapy*, 13, 13-25.
75. Laws, D. R. & Osborn, C. A. (1983). How to build and operate a behavioral laboratory to evaluate and treat sexual deviance. In J. G. Greer & I. R. Stuart (Eds.), *The sexual aggressor: Current perspectives on treatment* (pp. 293-335). New York, NY: Van Nostrand Reinhold.
76. Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse: A Journal of Research and Treatment*, 14(3), 207-223.
77. Looman, J., & Marshall, W. L. (2001). Phallometric assessments designed to detect arousal to children: The responses of rapists and child molesters. *Sexual Abuse: A Journal of Research and Treatment*, 13, 3-13.
78. Lykins, A. D., Cantor, J. M., Kuban, M. E., Blak, T., Dickey, R., Klassen, P. E., & Blanchard, R. (2010). The relation between peak response magnitudes and agreement in diagnoses obtained from two different phallometric tests for pedophilia. *Sexual Abuse: A Journal of Research and Treatment*, 22(1), 42-57.
79. Mahoney, J., & Strassberg, D. (1991). Voluntary control of male sexual arousal. *Archives of Sexual Behavior*, 20(1), 1-16.
80. Malamuth, N. & Malamuth, E. (1999). Integrating multiple levels of scientific analysis and the confluence model of sexual coercers. *Jurimetrics: Journal of Law and Science*, 39, 157-179.
81. Marshall, W. L. (1973). The modification of sexual fantasies: A combined treatment approach to the reduction of deviant sexual behavior. *Behavior Research & Therapy*, 11, 557-564.
82. Marshall, W. L. (1979). Satiation therapy: A procedure for reducing deviant sexual arousal. *Journal of Applied Behavior Analysis*, 12, 377-389.
83. Marshall, W. L. (2006). Clinical and research limitations in the use of phallometric testing with sexual offenders. *Sexual Offender Treatment*, 1(1), 1-25.
84. Marshall, W. L. (2007). Covert association: A case demonstration with a child molester. *Clinical Case Studies*, 6, 218-231.
85. Marshall, W. L. (2014). Phallometric assessment of sexual interest: An update. *Current Psychiatry Reports*, 16, 428-434.
86. Marshall, W. L. & Eccles, A. (1991). Issues in clinical practice with sex offenders. *Journal of Interpersonal Violence*, 6(1), 68-93.
87. Marshall, W. L. & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychological Review*, 20, 807-822.
88. Marshall, W. L. & Fernandez, Y. M. (2001). Sexual preferences: Are they useful in the assessment and treatment of sexual offenders? *Aggression and Violent Behavior*, 8, 131-143.
89. Marshall, W. L. & Fernandez, Y. M. (2003). *Phallometric testing with sexual offenders: Theory, research and practice*. Brandon, VT: Safer Society Press.
90. Marshall, W. L., Barbaree, H. E., & Butt, J. (1988). Sexual offenders against male children: Sexual preferences. *Behavior Research Therapy*, 26(5), 383-391.
91. Marshall, W. L., Laws, D. R., & Barbaree, H. E. (1990). *Handbook of sexual assault* (Eds.). New York and London: Plenum Press.
92. Martin, L. M & Potts, G. (2009). Impulsivity in decision-making: An event-related potential investigation. *Personality and Individual Differences*, 46, 303-308.
93. Masters, W., & Johnson, V. (1966). *Human sexual response*. New York, NY: Bantam Books.
94. McAnulty, R. D., & Adams, H. E. (1991). Voluntary control of penile tumescence: Effect of an incentive and a signal detection task. *Journal of Sex Research*, 28, 557-577.
95. McPhail, I. V., Hermann, C. A., Fernane, S., Fernandez, Y. M., Nunes, K. L., & Cantor, J. M. (2017). Validity of phallometric tests for sexual interests in children: A meta-analytic review.

Assessment. Manuscript accepted for publication.

96. Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts*, third edition. New York, NY: The Guilford Press.
97. Mental Hygiene Law, Chapter 27, Title B, Mental Health Act, Article 10.
98. Merdian, H. L. & Jones, D. T. (2011). Phallometric assessment of sexual arousal. In D. P. Boer, R. Eher, M. H. Miner, F. Pfafflin, & L. A. Craig, (Eds.), *International Perspectives on the Assessment and Treatment of Sexual Offenders: Theory, practice, and research* (pp. 141-169). New York, NY: John Wiley & Sons.
99. Michaud, P. & Proulx, J. (2009). Penile-response profiles of sexual aggressors during phallometric testing. *Sexual Abuse: A journal of Research and Treatment*, 21(3), 308-334.
100. Moffitt, T. E. (1993). Life-course-persistent and adolescence-limited antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674-701.
101. Mulhauser, K. R. W., Struthers, W. M., Hook, J., Pyykkonen, B. A., Womack, S. D., & MacDonald, M. (2014). Performance on the Iowa Gambling Task in a sample of hypersexual men. *Sexual Addictions & Compulsivity*, 21, 170-183.
102. Murphy, W. D., & Barbaree, H. E. (1987). *Assessments of sexual offenders by measures of erectile response: Psychometric properties and decision making*. National Institute of Mental Health. Washington, D.C.: U.S. Government Printing Office.
103. Murphy, W. D., & Barbaree, H. E. (1994). *Assessment of sex offenders by measures of erectile response: Psychometric properties and decision making*. Brandon, VT: Safer Society Press.
104. Murphy, L., Ranger, R., Federoff, J. P., Stewart, H, Dwyer, R. G., & Burke, W. (2015). Standardization of penile plethysmography testing in assessment of problematic sexual interests. *Journal of Sexual Medicine*, 12(9), 1853-1861.
105. Nordgren, L. F., van Harreveld, F., van der Pligt, J. (2009). The restraint bias: How the illustration of self-restraint promotes impulsive behavior. *Psychological Science*, 20, 1523-1528.
106. Odesheo, J. R. (2004). Of Penology and Perversity: The Use of Penile Plethysmography on Convicted Child Sex Offenders. *Temple Political & Civil Rights Law Review*, 14(1), 1-44.
107. O'Donohue, W. & Letourneau, E. (1992). The psychometric properties of the penile tumescence assessment of child molesters. *Journal of Psychopathology and Behavioral Assessment*, 14(2), 123-174.
108. Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, 6, 768-774.
109. Pfaus, J. G. (2007). Models of sexual motivation. In E. Janssen (Ed.), *The Psychophysiology of sex* (pp. 340-362), Bloomington, IN: Indiana University Press.
110. Proulx, J. (1989). Sexual preference assessment of sexual aggressors. *International Journal of Law and Psychiatry*, 6, 431-441.
111. Quinsey, V. L. & Bergersonn, S. G. (1976). Instructional control of penile circumference in assessment of sexual preference. *Behavior Therapy*, 7, 489-493.
112. Quinsey, V. L. & Carrigan, W. F. (1978). Penile responses to visual stimuli: Instructional control with and without auditory sexual fantasy correlates. *Criminal Justice and Behavior*, 5, 331-341.
113. Quinsey, V. L. & Laws, D. R. (1990). Validity of physiological measures of pedophilic sexual arousal in a sexual offender population: A critique of Hall, Proctor, and Nelson. *Journal of Counseling and Clinical Psychology*, 58(6), 886-888.
114. Quinsey, V. L., Steinman, C. M., Bergesen, S. G., Holmes, T. F. (1975). Penile circumference, skin conductance, and ranking responses of child molesters and 'normals' to sexual and nonsexual visual stimuli. *Behavioral Therapy*, 6(2), 213-219.
115. Richardson, 1969. *Mental imagery*. New York, NY: Springer.

116. Rosen, R. C. (1973). Suppression of penile tumescence by instrumental conditioning. *Psychosomatic Medicine*, 35(6), 509-514.
117. Rosen, R. C. & Keefe, F. J. (1978). The measurement of human penile tumescence. *Psychophysiology*, 15, 366-376.
118. Sachs, B. D. (2000). Contextual approaches to the physiology and classification of erectile function, erectile dysfunction, and sexual arousal. *Neuroscience and Biobehavioral Reviews*, 24, 541-560.
119. Sachs, B. D. (2007). A contextual definition of male sexual arousal. *Hormones and Behavior*, 51, 569-578.
120. Schuermann, B., Kathmann, N., Stiglmayr, C., Renneberg, B., & Endrass, T. (2011). Impaired decision making and feedback evaluation in borderline personality disorder. *Psychological Medicine*, 41, 1917-1927.
121. Seto, M. C. & Kuban, M. (1996). Criterion-related validity of a phallometric test for paraphilic rape and sadism. *Behavior Research Therapy*, 34(2), 175-183.
122. Seto, M. C., Lalumière, M. L., & Blanchard, R. (2000). The discriminative validity of a phallometric test for pedophilic interest among adolescent sex offenders against children. *Psychological Assessment*, 12(3), 319-327.
123. Simon, W. T & Schouten, P. G. (1991). Plethysmography in the assessment and treatment of sexual deviance: An overview. *Achieves of Sexual Behavior*, 20(1), 75-91.
124. Singer, B. & Totes, F. M. (1987). Sexual motivation. *Journal of Sex Research*, 4, 481-501.
125. Slavin, M. O. & Kriegman, D. (1992). The adaptive design of the human psyche: Psychoanalysis, evolutionary biology, and the therapeutic process. New York, NY: Guilford Publications.
126. Smith, R. E., Sarason, I. G., & Sarason, B. R. (1982). *Psychology: The frontiers of behavior*. New York, NY: Harper & Row, Publishers, Inc.
127. Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. I., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences*, 47, 385-395.
128. Tong, D. (2007). The penile plethysmograph, Abel assessment for Sexual Interest, and MSI-II: Are they speaking the same language? *The American Journal of Family Therapy*, 35, 187-202.
129. Vognsen, J. & Phenix, A. (2004). Antisocial Personality Disorder is not enough: A reply to Sreenivasan, Weinberger, and Garrick. *The Journal of the American Academy of Psychiatry and the Law*, 32, 440-442.
130. Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing Risk for Violence, Version 2*. British Columbia, Canada: Mental Health, Law, & Policy Institute.
131. Weinberger, L. E., Sreevivasan, S., Azizian, A., & Garrick, T. (2018). Linking mental disorder and risk in sexually violent person assessments. *Journal of the American Academy on Psychiatry and the Law*, 46, 63-70.
132. White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., Needles, D. J., & Stouthamer-Loeber, M. (1994). Measuring impulsivity and examining its relationship to delinquency. *Journal of Abnormal Psychology*, 103(2), 192-205.
133. Wilson, R. J. (1998). Psychophysiological signs of faking in the phallometric test. *Sexual Abuse: A Journal of Research and Treatment*, 10(2), 113-126.
134. Wilson, R. J. (2016). The use of phallometric testing in the diagnosis, treatment, and risk management of male adults who have sexually offended. In L. Craig & M. Rettenberger, (Eds.), *The Wiley handbook on the theories, assessment & treatment of sexual offending: Volume 2* (pp. 823-849). Chichester, UK: Wiley-Blackwell.
135. Wilson, R. J. & Miner, M. H. (2016). Measurement of male sexual arousal and internet using penile plethysmograph and viewing time. In D. R. Laws & W. O'Donohue (Eds.), *Treatment of sex offenders: Strengths and weaknesses in assessment and intervention* (pp. 107-131).

Cham, Switzerland: Springer International Publishing.

136. Winsmann, F. (2012). Assessing volitional impairment in sexually violent predator evaluations. *Sexual Offender Treatment*, 7(1), 1-14.
137. Winsmann, F. (2015, November). The psychological concept of volition. In F. Winsmann (Chair), *The third annual Boston symposium on psychology and the law: Volition and the law*. The Boston symposium on psychology and the law, Boston, MA.
138. Winters, J., Christoff, K. & Gorzalka, B. (2009). Conscious regulation of sexual arousal in men. *Journal of Sex Research*, 46(4), 330-343.
139. United States of America v. Walter Wooden, No. 12-7607 (4th Cir. 2018)
140. Wylie, K. R. & Eardley, I. (2006). Penile size and the 'small penis syndrome.' *BJU International*, 99, 1449-1455.
141. Zander, T. K. (2005). Civil commitment without psychosis: The law's reliance on the weakest links in psychodiagnosis. *Journal of Sexual Offender Civil Commitment: Science and the Law*, 1, 17-82.
142. Zuckermann, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin*, 75, 297-329.

Footnotes

¹The United States Court of Appeals for the Fourth Circuit, in its decision in *USA v. Wooden* (2018), favorably recognized Winsmann (2012) as the first article in the field that attempted to outline a protocol for assessing volitional impairment in sexual offender civil commitment cases.

²In the State of New York, both "serious difficulty in controlling behavior" (or "such conduct") and the higher standard of "such an inability to control behavior" are components of the civil management of sexual offenders. The civil management of sexual offenders is addressed in New York Mental Hygiene Law, Article 10 (Mental Hygiene Law, Chapter 27, Title B, Mental Health Act, Article 10). The law refers to management, instead of commitment, as individuals may either be managed in the community under supervision or managed via commitment where they are confined. New York bifurcates SVP proceedings under Article 10, where the fact-finder first determines if an individual is a sexual offender with a Mental Abnormality, which only requires a "serious difficulty in controlling behavior." The higher standard of "such an inability to control behavior" is required in New York in order to confine an individual rather than only supervise the individual in the community. This higher standard appears to be akin to the standard that was part of the U.S. Supreme Court decision in *Kansas v. Hendricks* (1997).

³It is also noteworthy that sexual arousal has been described as the sine qua non of the paraphilic disorders (First & Halon, 2008).

⁴Of note, the "dual control model" is not unlike the "inhibition hypothesis of sexual aggression" posited by Barbaree, Marshall, & Lanthier (1979). This hypothesis explains that sexual arousal involves excitation and inhibition and that sexual aggression likelihood increases when force or rejection of consent fail to inhibit sexual arousal. This hypothesis is an alternative to the sexual preference hypothesis, which states that if an individual is maximally aroused by an inappropriate stimulus, the eventual satisfaction will be greater than that resulting from an appropriate stimulus (Marshall, Laws, & Barbaree, 1990; O'Donohue & Letourneau, 1992).

⁵The phrase "full erection," hereinafter in this article, refers to the change in tumescence from flaccidity to full erection.

⁶From "Circumferential change scores in phallometric assessment: Normative data," by R. J. Howes (2003), *Sexual Abuse: A Journal of Research and Treatment*, 15(4), p. 370. Copyright (2003) Sage Publishing. Reprinted and adapted with permission.

⁷As another example, suppression PPGs are utilized after treatment at Coalinga State Hospital in the State of California, which houses male sexual offenders including those civilly committed, to examine the ability to suppress sexual arousal.

⁸The PPG discussed in this article is manufactured by Limestone Technologies, Inc. This is noted because different manufacturers utilize different technology.

⁹A basic outline of the administration of the PPG, and the suppression administration of the PPG, is provided, while use of the PPG requires training. This outline should not be considered an exhaustive or complete description of the proper administration of the PPG, or, by any means, a substitute for such training.

¹⁰Examiners should use professional judgment in determining whether more or less PPG trials are indicated to construct the VI-PPG. Examiners may also decide to administer a complete PPG as part of the VI-PPG and choose select trials to re-administer in the suppression condition. However, the initial administration needs to have a sufficient number of trials targeting the particular inappropriate arousal. Some stimulus sets limit a focused area to as few as one trial.

¹¹If only two of three trials reached at least 4.7mm of stretch in the unsuppressed condition, the average is based only on these two trials.

Author Note

Frederick Winsmann, Department of Psychiatry, Harvard Medical School, Cambridge Health Alliance. The author declares that he does not have financial interest in any product or device mentioned in this article, that he has no known conflict of interest, and this article does not contain any studies with human or animal subjects performed by the author.

Author address

Frederick Winsmann

262 Beacon Street

Boston, MA 02116

fwinsmann@frederickwinsmann.com