

Chapter 1

PENILE PLETHYSMOGRAPHY: STRENGTHS, LIMITATIONS, INNOVATIONS¹

D. RICHARD LAWS

Pacific Psychological Assessment Corporation, Victoria, BC, Canada

STRENGTHS

Penile Plethysmography as Technology

Dorland's Illustrated Medical Dictionary (1994) defines *plethysmography* as 'the recording of the changes in the size of a [body] part as modified by the circulation of the blood in it' (p. 1306). This is done by measuring electrical *impedance*, 'a technique for detecting blood volume changes in a part by measuring changes in electrical resistance'. Impedance changes are detected by a *strain gauge*, 'a technique for detecting blood volume changes in ... circumference employing a rubber tube filled with a conductive liquid; as the tube expands and contracts, the resistance in the fluid changes in proportion to the circumference'. A *plethysmograph* is 'an instrument for determining and registering variations in the volume of an organ' (*Dorland's Illustrated Medical Dictionary*, 1994, p. 1306, *passim*).

In medicine, plethysmography is primarily used for 'measuring changes in body volume, used especially in measuring pulmonary ventilation' as well as for 'measuring blood volume taking place in a single finger' (*Dorland's Illustrated Medical Dictionary*, 1994, p. 1306). This description shows that it is a small step from measuring blood volume in a single finger to measuring circumferential change in a penis. This impressive definition lends an aura of scientific respectability to penile plethysmography.

There are two general methods for measuring changes in penis size. The first, called the *volumetric method*, was developed by Freund in the former

¹This chapter is an adapted version of Laws, D.R., Penile Plethysmography: Will we ever get it right? In T. Ward, D.R. Laws and S.M. Hudson (Eds) *Sexual deviance: Issues and controversies*, 82–102. Copyright © 2003 Sage Publications, Inc.

8 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

Czechoslovakia. This method encloses the penis in a glass tube and measures changes in air volume in the tube as the penis expands and contracts (Freund, Sedlacek and Knob, 1965). Although this approach is highly sensitive and favoured by a small coterie of researchers, it has never seen wide use due to its expense and cumbersome nature.

The second approach, called the *circumferential method*, is virtually identical to the medical definition given above. A loop of silicone rubber tubing is filled with mercury or indium–gallium and plugged with electrodes attached to electronic circuitry that passes a weak current through the mercury (Bancroft, Jones and Pullan, 1966). This is called the *mercury gauge*. The transducer loop is fitted over the shaft of the penis. In a state of flaccidity, the resistance in the circuit is zeroed. Following this, any expansion of the penis will thin out the column of mercury and increase electrical resistance. These resistance changes are typically read out as a real-time tracing of the response. An alternative device (the *Barlow gauge*) was developed by Barlow, Becker, Leitenberg and Agras (1970). This method employs a mechanical strain gauge attached to the flat top of a thin band of surgical steel which is shaped like a ring and open on the lower side. It is placed on the penile shaft in the same way as the mercury transducer. When the penis expands, the strain gauge is slightly bent, which increases the resistance and permits the same kind of reading as the mercury gauge. The function of the two gauges is quite similar (Laws, 1977).

These are respectable scientific methods. If used properly, they should produce reliable and valid data. The manufacturers of penile plethysmographs provide extensive instructions for the use of their equipment. Some treatment centres such as Sand Ridge Secure Treatment Center in Wisconsin and HM Prison Service in the United Kingdom have developed highly detailed protocols for implementation and scoring of the procedure, but this is not typical. It is well recognized by researchers in the field that PGG equipment and procedures are often used in highly idiosyncratic ways, likely without continuous reference to user manuals. Despite these shortcomings, there resides a considerable confidence in this technology. ‘Machines don’t lie’, one of the author’s assistants once said. Since we know that penile plethysmography (PPG) evaluations can differ wildly from site to site, it matters little whether machines lie or not.

PPG as Procedure

Although PPG has the potential to be creatively used in a variety of ways, there are two typical procedures that have developed over the years.

Age and gender assessment

In this procedure, still images of males and females of various ages are presented to the client. There are typically five or six age categories with two to four exemplars per category. The images are presented for a predetermined period of time, usually 2 minutes. The client is instructed to allow himself to become sexually aroused if he finds the image sexually attractive. The resulting erection response values are

scored as millimetres off baseline, percentage of full erection, some kind of deviance index or transformed to ipsative z-scores. Means are computed for each category and the resulting scores per category are used to determine sexual interest by age and gender. This procedure has proven most useful with extrafamilial child molesters.

Sexual activity assessment

In this procedure, scenarios are constructed that describe various forms of consenting and non-consenting sexual activities and recorded on tape or CD. These can be constructed by the client but most often are prepared in a standard set by the researcher/clinician. The scenarios are usually 2–3 minutes in length. As with the age/gender assessment, the client is instructed to allow himself to become sexually aroused if he finds the script to be sexually attractive. This procedure is most often used with child molesters and rapists. The scripts contain different content to reflect the supposed sexual interest of these two groups.

The two procedures are directed primarily at child molesters and rapists because they make up the majority of clients assessed. Incest offenders are frequently exposed to the age/gender assessments but tend to respond like non-offenders. Specially prepared assessments using recorded scripts are sometimes prepared to evaluate clients such as exhibitionists, voyeurs or frotteurs.

PPG as Art

PPG is, in a sense, an art because there are essentially no universally agreed-upon standards for performing the procedure. For example, Howes (1995) conducted a survey of 48 plethysmographic assessment centres in 25 US states and 6 Canadian provinces. He was concerned that

although the technical adequacy of plethysmographic assessment is the subject of some disagreement, and validity studies are not entirely convincing, plethysmography had nonetheless been accepted as both a reasonably precise quantification of sexual arousal . . . and a diagnostic instrument about which there is every reason to be optimistic. . . . Perhaps the most substantial criticism of this procedure . . . is its apparent lack of standardization. (p. 14)

The agencies responding to Howes's questionnaire reported being in the plethysmographic testing business for an average of 5.5 years ($R = 6$ mo – 25 yr). Technicians performing the assessments reported an average of 3.4 years of experience ($R = 6$ mo – 25 yr). Forty-two per cent of the technicians had been doing the assessments for 2 years or less. Formal training in the procedure was 1 week or less for 76% of the technicians, and 18% had received no training at all. Admittedly, the extent to which Howes's sample of 48 is representative of North American practice is unknown. The data he reported are appalling. The author's several decades of experience with the procedure and with a variety of assessment centres suggest to me that he is right.

10 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

O'Donohue and Letourneau (1992) also noted that 'there does not appear to be a standardized penile tumescence assessment, but rather there is a family of procedures which share some common aims and features' (p. 126). They listed the following potential sources of procedural variation. These referred to the assessment of child molesters but would apply equally well to other offender groups:

1. Type of strain gauge used (mechanical, mercury) and transducer placement.
2. Type of stimuli used (audio recordings, slides, videotapes).
3. Content of stimuli used (differences in models).
4. Duration of stimulus presentation (2 s to >4 min).
5. Length of interstimulus (detumescence) interval (fixed time vs. return to baseline).
6. Nature of stimulus categories sampled (Tanner criteria vs. age scales).
7. Number of categories and of stimuli used for each category.
8. Instructions to subjects (imagine sexual behaviour with target vs. no instructions).
9. Whether a warm-up was used and number of assessment sessions.
10. Type of recording instrumentation used (computer-generated graphs vs. strip chart recorder).
11. Whether strain gauge calibration was used to correct for any non-linear characteristics of recording.
12. Data sampling rate (every 5 s vs. every 1 min).
13. Whether methods were used to assess for faking.
14. Gender and other characteristics of the evaluator.
15. Type of data transformation (z-score vs. deviance index).
16. Characteristics of the laboratory (degree of privacy).
17. Type of sample and setting (outpatient, prison).

These appear to be a formidable set of requirements but actually they are not. They are exactly the problems that one encounters every day in performing these assessments. Over the years, the author has visited many assessment centres and spoken to the persons in charge. He has stressed the absolute necessity of standardization to them. He has routinely received responses such as, 'We've always done it this way', or 'We do what works for us', or 'We've tested hundreds of people using this method. Why should we change?' For 8 years, first as president and then as a member of the Executive Board of the Association for the Treatment of Sexual Abusers (ATSA), the largest umbrella organization in the world for workers in this field, the author repeatedly stressed the need for proper training in psychophysiological assessment.

However, ATSA made an effort. In their *Practice Standards and Guidelines for Members of the Association for the Treatment of Sexual Abusers* (ATSA, 1997), they included an appendix (pp. 40–43) dealing with the plethysmographic examination. It should be emphasized that this appendix is *not* intended as a training manual. It is simply a set of recommendations and guidelines. It includes, in part, (a) requirements for training, (b) appropriate client groups, (c) screening of clients, (d) informed consent, (e) appropriate stimulus sets, (f) legislation regarding use of erotic stimuli, (g) stimulus material, visual and audio, (h) documentation of assessment data and

(i) data scoring and interpretation. This set of very general guidelines meets some but not all of the issues raised by O'Donohue and Letourneau (1992). The extent to which practitioners followed the recommendations is unknown. The latest revision of this manual is equally non-specific (ATSA, 2001).

In recent years two highly detailed procedure manuals have been prepared, one by the Sand Ridge Secure Treatment Center in Wisconsin (D.M. Thornton, personal communication, 5 April 2007) and the other by HM Prison Service in the United Kingdom (HM Prison Service, 2005). The author was asked to peer review the latter manual. It describes, in considerable detail, the exact elements of the assessment procedure as well as the clinical protocol that must accompany it. It is a hard read but everything is there. It embodies precisely what many researchers have been recommending for 20 years.

There is also a national effort underway in the United Kingdom. D. Perkins (personal communication, 11 April 2007) has described this process:

There is now a national PPG forum in the UK, bringing together practitioners from prisons and forensic mental health services. The group has been asked by the BPS [British Psychological Society] to update the current BPS PPG guidelines, which we are currently doing. Different sites are using slightly different procedures, stimulus sets etc., but the group is helping establish some common practices. It is also a useful forum for considering complex or controversial PPG referrals/results, and this is proving valuable as a peer reference group. We are trying to set up some comparison of visual and auditory stimuli along with other variables. We have a project ongoing on the use of film material in ways that are ethical and legal (including [sic] copywrite), and a couple of us are looking at visual scan technology linked to PPGs.

To date, there has been one major attempt in North America to standardize the age and gender assessment for child molesters. In 1987, a group of senior researchers in penile plethysmography met at the National Institute of Mental Health in Rockville, Maryland. The purpose of the meeting was to define a standard protocol for stimulus type and procedure in that assessment (Abel, Becker, Card, Cunningham-Rathner, Farrall, Jensen, Laws, Murphy, Osborn, Quinsey and Wormith, 1989). This was to be called the Multisite Assessment Study and was to be carried out at five sites in the United States and three in Canada. Only one Canadian site (Laws, Gulayets and Frenzel, 1995) completed the study. The protocol (see Laws *et al.*, 1995, pp. 48–52) specified (a) characteristics of the research participant, (b) characteristics of the individual stimulus slides, (c) characteristics of the slide set, (d) calibration of equipment, (e) details of the assessment procedure, (f) informed consent, (g) data reduction and (h) statistical analysis of data. This protocol attempted to address most of the concerns raised by O'Donohue and Letourneau (1992). The study by Laws *et al.* (1995) reported on only 30 participants. Had the other sites reported that many subjects, we would have had the beginnings of an international database. No further attempt has been made to complete this project.

Murphy and Barbaree (1994) observed that most of the published studies in the literature have originated in the laboratories of a small number of senior investigators. Students and colleagues of these researchers continue to publish their

12 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

findings. At least for this small coterie of investigators one may reasonably surmise that there is a rough consistency in their procedures and manner of reporting data. But what of the vast majority of clinicians/researchers who were not part of this more experienced group, who learned to do PPG by listening to conference presentations or reading the publications of this group? Therein lies the major continuing problem in standardization.

In the past few years, there have been discouraging as well as encouraging developments. In 2005, a meeting on standardization was held at the annual ATSA conference. About 50 people attended a meeting. The author had reviewed the HM Prison Service procedure manual and protocol a few months earlier and presented the basic idea of a potential nationwide protocol. The organizer of the meeting had set up a web site (phallometrics.com) and people were encouraged to blog on this site. There seemed to be some general interest and then the questions started: 'Why should I devote precious time to this procedure when the one I have works perfectly well?' And that was the end of it. Hardly anyone contributed to the web site and it is now defunct. On the other hand, the remarks by Perkins cited above are encouraging. The United Kingdom is a small country, the persons doing this work are all known to one another, and the likelihood that a reasonably coordinated system will develop is high.

Related to these issues is a multisite study proposed by Marshall, Fernandez, Marshall and Mann (2001). An investigation was proposed that would continue over a 2-year period at multiple sites in Canada and the United Kingdom. The study would compare responses of extrafamilial child molesters, incest offenders, rapists and normals. These authors estimated that the final data set would exceed 1000 participants. The stimulus sets used would be recorded descriptions of sexual activity between adult males and children and of sexual violence against adult women. The sets would be updated versions of stimuli previously tested and reported by Quinsey and his colleagues (Quinsey and Chaplin, 1988; Quinsey, Chaplin and Varney, 1981). The project was proposed to proceed in four phases: standardization, data collection, psychometric analyses and establishment of normative data. All sites would use the same assessment protocol, the same PPG equipment, the same stimulus sets and the same data-reporting format. Three studies were proposed. The first would focus upon internal consistency, the second on test-retest reliability and the third on criterion validity, examining differences between the four groups. The distribution of scores for the normal groups would be determined. Percentile scores for those distributions would be established in order to compare the offenders' responses to them.

The main deficiency in the proposal is the absence of an age/gender assessment. The age/gender assessment using visual stimuli has been the classical plethysmographic procedure with child molesters for decades. Its absence here is regrettable.

Dean (C. Dean, personal communication, 12 April 2007) has provided an update on this research programme. On the noted date, 7 sites were using the standardized PPG protocol. Three sites have been using them for over 2 years. By the end of 2007, it is expected that all 10 sites proposed will be fully equipped and running. Across the sites currently operating, over 100 PPGs have been completed, the majority being pre- and post-intervention. The main addition to the procedure has been the development of an *Interpretation Manual*. Dean stated that this manual 'clearly

documents procedures for interpreting and reporting data, including our decision-making criteria regarding when responses should be considered interpretable and indicative of offence-related sexual arousal'.

In the author's judgment, the proposal of Marshall *et al.* (2001) represents a breakthrough study that has been too long in arriving. It will undoubtedly solve many of the problems that have plagued PPG over the years. However, we may reasonably ask: Will this information be accepted beyond the proposed study sites? If it is simply reported in the psychological literature, the answer is probably 'no'. On the other hand, if the researchers package the protocols, the stimulus sets and the normative data, sell them or even give them away, there is an excellent chance that this model could be adopted. We must hope for that result.

Summary on Strengths

The preceding section has included some harsh comments about PPG. Although the method is intrusive, invasive of privacy and very time consuming, it works well if implemented in a relatively consistent fashion. Therein are the problems which compromise what otherwise should be viewed as a scientifically respectable psychophysiological assessment procedure. PPG has been used to evaluate sex offenders since the mid-1960s (Bancroft *et al.*, 1966). It has only been very recently, 40 years later, that we are seeing the emergence of standardized procedures and explicit protocols for the procedure. This should never have happened. The many problems have been detailed for years. It is time to fix them.

Does PPG have strengths in clinical and research use? It most certainly does. However, those strengths reside where they have always resided, in the age/gender and sexual activity assessments. Despite dramatic improvements in equipment, stimulus presentation and scoring of data, these procedures continue to be implemented in the same manner as 30 years ago. The introduction of highly specific procedure manuals is a very welcome addition.

LIMITATIONS

What about Construct Validity?

Vogt (1993) defined a 'construct' as 'something that exists theoretically but is not directly observable. . . . A theoretical (not operational) definition in which concepts are defined in terms of other concepts' (p. 44). 'Deviant sexual arousal' is such a construct. Construct validity, said Vogt, refers to 'The extent to which variables accurately measure the constructs of interest. . . . Do the operations really get at the things you are trying to measure?' (p. 44). In terms of our present concern, is the phenomenon that PPG measures – penile erection – a valid measure of deviant sexual arousal (i.e. deviant sexual interest and preference)? Opinion has been divided on this issue for decades. Many clinicians and researchers believe that PPG is, in fact, a valid measure of deviant sexual interest.

14 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

This point of view has not gone unchallenged. For example, O'Donohue and Letourneau (1992) noted that

[p]enile measurement is often used to gather information so that inferences can be made concerning naturalistic behavior (e.g., Does this individual prefer children over adults as sexual partners?). However, this type of measurement as currently practiced uses neither naturalistic stimuli nor naturalistic responses. . . . Penile tumescence measurement is more directly an assessment of penile response to erotica than a measure of actual sexual behavior in naturalistic situations.

As with any analogue assessment, the nature of this methodology may miss many critical elements that are relevant to sexual behavior and sexual offending (e.g., affection, fear, tactile and olfactory clues [*sic*]). Moreover, the only sexual behavior measured is penile responding, and other relevant behavior such as verbalizations, approach behavior, touching, cognitions, etc. are usually ignored. . . . Thus, penile measurement techniques do not involve a direct sampling of the domain of interest, but, rather, involve an indirect, analogue approach. More bluntly, pen deflection is not directly [*sic*] sexual preference and viewing a slide is not a naturalistic potential sexual interaction. (pp. 162–163)

An opposing position was advanced by Barlow in 1977:

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. (Cited in Laws and Osborn, 1983, p. 294)

Laws and Osborn (1983, p. 295) acknowledge that the problem of ecological validity is an important one but agree with Barlow that the erection response is measured because it is the one behaviour in the chain that *can* be (more or less) objectively measured.

Murphy and Barbaree (1994) reported that early research in construct validity centred on comparing subjects' reports of their level of sexual arousal with the measured values. Some researchers (e.g. Abel, Blanchard, Murphy, Becker and Djenderedjian, 1981; Wincze, Wenditti, Barlow and Mavissakalian, 1980) found high correlations among these variables. Murphy and Barbaree (1994) also noted that competing stimuli and demand characteristics of the situation could affect self-report of arousal. When some of these competing variables were taken into account, correlations between measured arousal and self-report were quite low (e.g. Farkas, Sine and Evans, 1979).

There has been a movement towards investigating construct validity in terms of convergent validity. This more straightforward approach compares PPG measures with other indicators of deviant and non-deviant sexual interest. For example, Day, Miner, Sturgeon and Murphy (1989) reported a classification study in which they compared PPG responses to slides, audiotapes and videotapes with a self-report

measure that was constructed from the item pool of the Multiphasic Sex Inventory (MSI; Nichols and Molinder, 1984). The PPG measures correctly classified 82% of the offenders by sex of victim and 74% by both victim gender and use of force. The MSI measures, on the other hand, correctly classified 86% by sex of victim and 85% by use of violence. The combination of PPG and self-report was not tested.

Laws, Hanson, Osborn and Greenbaum (2000) reported a similar study. Child molesters completed a self-report card-sort measure of sexual interest and PPG responses were obtained from slides and audiotapes. All three measures of paedophilic interest significantly differentiated boy-object child molesters from girl-object offenders. The card sort measure showed the greatest classification accuracy and was the only measure to improve accuracy, once the other two modalities were considered. Taken together, all three modalities provided classification accuracy of 91.7%, greater than any single measure.

The investigations of Day *et al.* (1989) and Laws *et al.* (2000) appear to provide evidence for construct validity, that is, there is such a state as 'deviant sexual arousal'. We must ask, however, whether penile responses to a putatively erotic stimulus, a slide or a recorded description of sexual behaviour and endorsement of inventory items that purport to be descriptions of deviant behaviour are in fact measures of sexual interest, not to say preference. Some would say that they are not. It could be argued that acceptance of these data as measures of interest and preference are more an act of faith than an empirical demonstration of fact. To be sure, the data supporting construct validity are encouraging but they are not definitive.

Is PPG a Genuine Test?

O'Donohue and Letourneau (1992) have noted that the PPG procedure is not equivalent to typical standardized psychological tests such as the WAIS-R, which uses an invariant administration protocol. They ask whether PPG is a *norm-referenced* test, where a client's scores can be compared to established norms to determine if his score is 'deviant', or whether it is a *criterion-referenced* test, where the purpose is to indicate whether the subject responds to some established treatment criterion such as a 'normal' as opposed to a 'deviant' sexual response. As indicated previously, most PPG practice is so cluttered with procedural inconsistencies and the failure to establish a standard protocol that it could not, in its present state, come close to being termed a norm-referenced test. The cited proposal of Marshall *et al.* (2001) is a move towards resolving some of these issues. At present, the author is aware of only one attempt to produce normative PPG data. Howes (2001) collected data on 724 subjects from nine sites. He was able to establish norms for the interpretation of low arousal scores in cases where full erection could not be obtained. In treatment, but probably only in behaviour therapy, does PPG approach being a criterion-referenced test. In this application, the aim of the treatment is to decrease deviant arousal and increase non-deviant arousal. These changes are easily observable in behaviour therapy data. In other situations, the magnitude of the response becomes the criterion. If a penile response to a supposed deviant stimulus is greater than a response to a non-deviant stimulus, that is considered a criterion response.

16 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

Another possibility, said O'Donohue and Letourneau (1992), is that PPG is not really a test in the typical meaning of that term but is rather a *direct observation* of behaviour. This is a staunchly behaviourist point of view that would argue that sexual response is a phenomenon worth studying in its own right. Examples of this approach can be seen in the work of Abel (1979), Abel, Blanchard, Barlow and Mavissakalian (1975) and Laws (1984). In these investigations, differing levels of ongoing erection responding were compared to the content of concurrently presented audiotaped scripts.

Can PPG Scores Predict Reoffence?

Opinion and empirical research are divided on this question. There have been reports over the past decades that suggest that PPG scores are (usually weakly) related to sexual recidivism (e.g. Barbaree and Marshall, 1988; Quinsey, Chaplin and Carrigan, 1980; Rice, Quinsey and Harris, 1991). In the early 1990s, PPG use lay somewhat dormant due to concerns about the use of stimuli depicting nude children or violent and degrading acts. In 1998, Hanson and Bussière published a large meta-analysis that demonstrated that a 'paedophile index' (deviant responses divided by non-deviant responses) was a robust predictor of recidivism. Historically, this was important. The publication coincided with a continuing intense interest in meta-analysis and a surge of confidence regarding the use of actuarial assessment to predict sexual recidivism. In short, it made PPG respectable again.

The only example of this use of PPG in actuarial assessment may be seen in the *Sex Offender Risk Appraisal Guide* (SORAG) (Quinsey, Harris, Rice and Cormier, 1998, pp. 241–243). The SORAG is a 14-item algorithm based solely upon static risk factors. Item 13 is *Phallometric test results*. There are only two choices for this item:

All indicate non-deviant sexual preferences = –1
Any test indicates deviant sexual preferences = +1

Note the words *any test*. Does this refer to a single stimulus or a group of stimuli? It seems to say that *any* erection response to a deviant stimulus considered 'significant' by the evaluator adds one point to the total score. Because we do not know what a clinically significant response is, confidence in this item is diminished. Because PPG is not widely used in the typical outpatient forensic evaluation, it seems unlikely that it will be incorporated in future actuarial instruments.

There is a larger problem here. In the United States, some 'sexually violent predator' (SVP) programmes use PPG. When administrators of these programmes are asked why they are using such an assessment procedure on people who have been determined to be so dangerous that they might have to be permanently incarcerated, they are likely to reply that it is to determine 'treatment needs'. The result of this practice could be that, if deviant arousal does not decrease as a result of that treatment, it is a small step to cite Hanson and Bussière (1998) or, more recently, Hanson and Morton-Bourgon (2004) to assert that this is evidence that the person is still too dangerous to release. This is a highly inappropriate use of PPG data (see below).

Does PPG predict reoffence? Researchers citing earlier work by Quinsey *et al.* (1980) and, more recently, Hanson and Bussière (1998) or Hanson and Morton-Bourgon (2004) might argue that it does. Others will argue that simply because deviant sexual arousal has been observed in some recidivists hardly makes it a predictor. How those data are used requires rather careful consideration of many other variables operating in the individual's life which may serve as protective factors to decrease danger.

Legal Challenges to PPG

Lawyers for the defence often argue that their clients are being abused by legal, psychiatric or psychological procedures when PPG testing is ordered. They may argue that the procedure is humiliating to their client, that he is being forced to participate in what is clearly a degradation ritual. It is true that the procedure is highly invasive of personal privacy. It is equally true that the majority of technicians go to considerable lengths to ensure the comfort and well-being of their testees. But sex offenders are not attractive plaintiffs, and this argument may be given little weight.

There are growing legal challenges to PPG based on civil law. Consider the following hypothetical case (adapted from Rulo, 1999, p. 1):

John Smith is a defendant who has just pleaded guilty to several counts of child molestation. John is awaiting his sentencing hearing when he is told that the court has ordered him to take a test. John is told that in the administration of this test a device will be attached . . . to his penis, to monitor its response to various graphic images to which he will be exposed. The results of this test, John is told, will be used as evidence in his sentencing hearing to establish his current status as a sexual deviant and later as a condition of his release, to make predictions about his future as a sexual deviant and to monitor his rehabilitative progress.

The prosecution will argue that PPG is a well-established procedure for evaluating sexual deviants that has been in constant use for over 30 years and is supported by hundreds of publications in the professional literature. John's lawyer may counter-argue that existing precedents in civil law have determined that PPG is an unreliable and invalid procedure, totally lacking in standardization both for administration and interpretation and therefore is inadmissible as evidence in court. . . . John's lawyer is highly likely to win this point.

From 1923 to 1993, the standard for admissibility of scientific evidence in court was the Frye test. The Frye standard is that a scientific finding or practice be generally accepted in the scientific community from which it comes. Frye states:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the

18 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

decision is made *must be sufficiently established to have gained general acceptance in the field in which it belongs* [emphasis added]. (*Frye v. US*, 293 F. 1013 [D.C. Cir. 1923])

It is obvious that plethysmographic evaluation of sex offenders is a very small portion of behavioural science inquiry. Even within that narrow band of interest, PPG has never been fully accepted. Some of the reasons for this have been noted above. PPG has never met the stipulation of the Frye test.

A new, more comprehensive standard of admissibility was set in 1993 (*Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 113 S. Ct. 2786, 1125 L.Ed.2d 469 [1993]). Smith (1998) noted that

Frye mandated that scientific evidence was admissible only if 'generally accepted' by the scientific community. ... In 1993 the United States Supreme Court adopted a somewhat more flexible, factor-based approach to the admission of scientific evidence in *Daubert v. Merrell Dow Pharmaceuticals*. ... Where the *Frye* test rigidly adhered to a rule requiring 'general acceptance' ... the *Daubert* standard focuses on the 'reliability' and 'fit' (relevance) of the evidence. (pp. 2-3)

Smith (1998) reviewed much of the legal literature surrounding the admissibility of PPG evidence in court, an account too lengthy to summarize here. The main factors that emerged from *Daubert* are the following:

1. Has the technique been tested?
2. Has the technique been subject to peer review and publication?
3. What is known of the potential rate of error?
4. Do standards exist for the control of the technique's operation?
5. Has the technique been generally accepted within the relevant scientific community?

PPG meets the standard for points 1 and 2, although the results are fraught with inconsistency. The technique has been tested thousands of times but with highly variable results depending upon what 'rules' the examiner is following. It has been peer reviewed and published hundreds of times. Opinions of reviewers vary widely and are rarely unanimous. It is with the remaining points that PPG encounters the most difficulty.

The absolute rate of error is unknown (point 3). The potential for error is large. Smith (1998) stated, 'Penile plethysmography cannot meet *Daubert's* standard of validity or relevance tests because the test results are not generally accepted, are not sufficiently accurate, the test is subject to faking and voluntary control by test subjects' (pp. 5-6).

To obtain a flavour of the error rate in the procedure, one need only consult primarily favourable reviews by Murphy and Barbaree (1994) and O'Donohue and Letourneau (1992) and the more negative ones by Marshall and Fernandez (2000, 2001, 2003). Taken as a whole, it would appear that the error rate is substantial.

The final point in *Daubert* is a reiteration of *Frye*: Is PPG generally accepted in the relevant scientific community? PPG does not meet this standard for the following reasons:

1. Adequate standards for the administration of the procedure do not exist (Barker and Howell, 1992; Howes, 1995; Simon and Shouten, 1991). There are no centres where technicians can be trained. The first formal attempt at standardization, the Multisite Assessment Study (Abel *et al.*, 1989; Laws *et al.*, 1995), was never completed. As noted above that several attempts at standardization are currently being undertaken.
2. O'Donohue and Letourneau (1992, p. 126) listed 17 potential sources of variation in the procedure. Some of these are present in all testing centres.
3. Marshall and Fernandez (2000, p. 813) stated that the PPG procedure appeared to show reasonable but not impressive internal consistency. Test-retest reliability has not been established.
4. Marshall and Fernandez (2003) stated that there is no strong evidence for the criterion validity of the PPG test. This refers to the ability of the test to differentiate sex offenders from one another as well as from non-offenders. They found that the agreement across studies was much greater for child molesters than for rapists. Further, they stated that the test appears to have limited value if rapists and exhibitionists appear normal at assessment and only those extrafamilial child molesters who admit to being deviant show deviant arousal.
5. Laws and Rubin (1969) demonstrated that it was quite easy for males to suppress their erection responses. Laws and Holmen (1978) later demonstrated that, depending upon the type of instructions provided to a single client, he could produce a credible response or suppress the response at will, even while speaking a fantasy unrelated to the stimulus displayed. Although it may be difficult for some to produce a response in the presence of a non-preferred stimulus, a profile of no responding is uninterpretable.

Within the 'relevant scientific community' that routinely uses PPG, none of the preceding will have any impact. PPG will live on, bolstered by the findings of Hanson and Bussière (1998) and Hanson and Morton-Bourgon (2004) and the perceived absolute need to assess and predict risk.

Inappropriate Uses of PPG

In the conclusion of their monograph, Murphy and Barbaree (1994, pp. 84–85) consider the appropriate and inappropriate uses of PPG. Apart from the numerous classification studies seen in the literature, the most appropriate *clinical* application of the technique is the use of erection responses to indicate the need to target deviant sexual arousal for treatment and to monitor the effectiveness of that treatment. In the punitive contemporary social climate, that is a very small part of the bigger picture.

20 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

There are three major inappropriate applications of the procedure:

1. Use of erection responses to determine or make statements about whether someone has committed a specific sexual offence or whether someone 'fits the profile' of a sex offender.
2. Use of erection responses as a sole criterion to decide someone's release from custody or from a treatment programme.
3. Use of erection responses to screen general populations in search of potential sex offenders.

First, there is no such thing as a 'profile' of any variety of sex offender. The second point is the most egregious use of the technique. Mentioned above is the persisting problem that the meta-analyses of Hanson and Bussière (1998) and Hanson and Morton-Bourgon (2004) have introduced, linking deviant sexual arousal to recidivism. This presents a temptation in both custodial and community programmes to use sexual arousal as a key decision-making issue. At the moment, the third point may seem far-fetched. However, for example, given the concern over clerical abuse of children, it is not inconceivable that PPG could one day be introduced in seminaries where these individuals train, or made a prerequisite for employment to work with children, or made a prerequisite to obtain certain types of security clearance. There are many possibilities for improper use and all of them are grim.

Murphy and Barbaree (1994, p. 85) make two additional points that are worth mentioning:

1. Erection measures should always be used with other data. These include psychometric evaluations as well as clinical interview data, police reports and victim statements. They should never stand as a single mode of assessment.
2. Offenders are frequently accused of faking when they fail to respond in a PPG assessment. It should be remembered that failure to respond may occur for a variety of reasons, only one of which is faking. Data that show lack of response or very low responding should be reported as exactly that and viewed as uninterpretable.

Conclusions on Limitations

PPG was introduced as a clinical tool with sex offenders in the mid-1960s. It has actually changed very little since then. From the 1960s to the 1980s most laboratories used very large equipment. In those days, the author's laboratory used a 6-channel Beckman Dynograph which could record other events as well as arousal, with all electronic controls situated in a large power rack. It looked impressive but the functions were quite simple. The only thing that is different today is that we have much smaller and much superior electronics, stimulus presentation and data management by computer. These are artefacts of technological change, nothing more. The basic procedure is what it has always been and is still subject to all the same shortcomings. It is a ponderous procedure that eats up valuable clinical time. And, in the end, the actual yield of new information in the data is typically quite small. PPG often tells us what we already know.

Of one thing we can be certain. Whether empirical evidence supports its continued use or not, whether its procedural faults are remedied or not, PPG will survive. We have wasted a great deal of time on a procedure that has given us too little in return. Things could always have been different. The effort to make them different would require a degree of procedural cooperation among clinicians and researchers that has never been a trademark of this field.

INNOVATIONS

In the 42 years that PPG has been used to evaluate sex offenders (1965–2007) there have been very few innovations reported. As mentioned previously, there have been significant technological improvements in the procedures, but there has been little creative use of PPG in research or clinical applications. Importantly, when innovations have appeared, they have rarely been followed up. This section reports some of them.

PPG + Polygraphy

Considering the interest in polygraphy that has developed over the past decades, this seems a natural extension of the use of PPG. Thornton (D.M. Thornton, personal communication, 5 April 2007) has described the integration of the two procedures. The PPG portion is presented in two parts. The first, called the ‘non-suppression PPG’, is the standard procedure in which the client is instructed to attend to the stimulus and allow himself to become sexually aroused with no attempt to control the response. The second, called the ‘enhanced non-suppression PPG’, is the same except that the client is asked a series of questions about the stimulus 30 seconds after it has terminated. The purpose, says Thornton, ‘is to encourage him to process it more deeply’. Then, in the following days, the client undergoes a polygraph examination which focuses upon compliance with ‘a detailed list of dos and don’ts, and asking more generally whether he has deliberately tried to distort the results’.

Thornton states that, in addition to examining penile changes, clinicians also look at respiration and skin conductance changes. These are compared with the results of the polygraph examination. Then, all of this is put into the context of the client’s known history and what he has been saying in group meetings about his sexual interests.

The procedure of Thornton *et al.* goes well beyond what we have recognized over the years as the standard PPG assessment.

Evaluating Stimulus Content

One of the criticisms justifiably levelled at PPG is that one is never certain about what aspect of the stimulus, visual or auditory, the individual is responding. If the stimulus is judged to be deviant in nature and the individual responds at some criterion level, that response is labelled sexually deviant. One can, of course, ask

22 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

the subject: 'Why did you respond to that? What did you find attractive about it?' He may provide a straightforward and credible answer and he may not. In an effort to circumvent this problem, attempts have been made to match response patterns to stimulus content.

With regard to audio stimuli, Laws (1984, p. 132) observed:

Unlike visual stimuli, audiotaped descriptions control the erection response rather precisely. Subjects are instructed to attend to the spoken text and not to try to anticipate the described action or relate it to their own experience. 'If you feel yourself becoming sexually aroused', they are told, 'let that happen'. The result, generally speaking, is a highly variable response record, accelerating to subjectively erotic portions, decelerating to nonerotic portions. And so we find ... the subjective erotic value of a wide variety of sexual activities in a particular individual.

Abel *et al.* (1975) were the first to examine how audiotaped segments could control the erection response. Using a fantasy produced by a proclaimed shoe fetishist they were able to show that segments describing a woman's foot produced substantially greater arousal than segments describing only shoes.

Abel (1979) matched the text of a 120-second audiotape of a frottage offence against a PPG tracing of response to the tape. Broken into 6-second blocks it was possible to see exactly which portions of the text produced the greatest arousal. The subject was incrementally aroused by the whole episode. The greatest uptick in arousal was evident in the final 12 seconds where the text described the *sine qua non* of frottage – a victim trapped in the midst of uncaring, unobserving people while the frotteur calmly performed his offence. The offender even obliged by ejaculating at the finale, confirming the diagnosis. This admittedly is a dramatic example but shows the potential power of the procedure.

Laws (1984) replicated Abel's findings with a rapist, two bisexual paedophiles, two exhibitionists and a sexual sadist. The purpose of Laws' research was to demonstrate that offenders then believed to be not particularly dangerous (paedophiles, exhibitionists) were highly sexually aroused to explicit descriptions of sexual violence.

The Kurt Freund Phallometric Laboratory

This description is included here because the Kurt Freund unit is one of a kind. Wilson (R.J. Wilson, personal communication, March 2007) described the early set-up of the laboratory. There were two unique features at that time.

The first feature was the exclusive use of the volumetric plethysmograph (Freund *et al.*, 1965) described above. This is the only institutional laboratory using this device and thousands of clients have been tested with it. The author is aware of only two other private practitioners who use it.

The second feature was the use of a standardized set of 16-mm films as stimuli for assessment. These were initially developed by Freund in Czechoslovakia and abandoned when he resettled in Canada in 1969 at the Clarke Institute of Psychiatry

in Toronto. The set was reproduced in 1970–1972. In each brief stimulus presentation, the subject is nude, standing in front of a theatre curtain. He or she slowly approaches the camera. The only thing that changes is the model. The background and motion are identical and there is no extraneous influence due to context or activity. This set was revolutionary for its time as most other laboratories were using slide stimuli. This procedure was in place for decades.

For the past decade, the procedures at the Kurt Freund Laboratory have changed somewhat in that the motion picture films are no longer used (M. Kuban, personal communication, 9 May 2007). Kuban described the current procedures:

In the current testing for gender and age preference, audio-taped narratives are presented through headphones and accompanied by slides. Three projectors display the images simultaneously on three adjacent slide screens approximately two to three metres in front of the patient. In each display is an image of a nude model from each of three angles: full frontal view, full back view and close-up genital view. In the audio, there are seven categories of narratives, which describe mild sexual interactions with prepubescent girls, pubescent girls, adult women, prepubescent boys, pubescent boys, adult men, and also non-sexual neutral stimuli. The accompanying slides show nude models corresponding in age and sex to the topic of the narratives, and landscape slides accompany the neutral stories. The test stimuli are presented in discrete trials, each 54 seconds in duration, with between trial intervals for as long as required for penile volume to return to baseline. The full test consists of four blocks of seven trials, with each block including one trial of each type in a fixed pseudo-random order. The full duration of this test is about an hour. There are, however, other audio-only tests for sexual coercion, masochistic interests, or cross-gender (autogynephilia), which have been employed in slightly differing formats than the gender/age test.

The procedure described by Kuban is unique for PPG assessment in the use of three screens to present visual stimuli accompanied by audio narratives appropriate to the visual material shown. Some investigators would argue that this represents confounding of stimuli. However, Kuban reported that ‘less than 5% of our tests produce responses below those required for a valid test’.

Fading

In the 1970s, the use of stimulus fading was an attempt to alter sexual preference by positive rather than aversive conditioning. The following examples illustrate two of the methods. Only portions of the procedures are described.

Barlow and Agras (1973) attempted to treat three male homosexuals. Using adult male and female slide stimuli, they positioned two slide projectors so that the images overlapped on the screen. Using a variable voltage transformer they were able to manually dim or brighten each image. There was a series of 16 equal steps from 0% illumination to 100%. Subjects wore the Barlow strain gauge (Barlow *et al.*, 1970) to assess penile circumference. Subjects experienced a 2-min presentation with a male image brightened to 100% and the female image to 0%. If the

24 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

subject produced an erectile response equal to 75% of his maximum, he advanced to the next step (6% female, 94% male). This stepping process continued as long as his response remained above criterion. If he failed to meet the criterion, the image was displayed in repeated trials until he did. There were six trials per session. Generalization was evaluated in separate sessions which measured response to single-slide presentations. Although homosexual arousal remained high across sessions, Barlow and Agras reported about a 25–30% increase in heterosexual arousal in the post-fading generalization sessions.

Laws and Pawlowski (1974) presented an automated version of the fading procedure to two paedophiles. Their approach resembled that of Barlow and Agras with these exceptions: (1) they attempted to alter sexual interest *within* a sexual orientation, (2) a continuous session rather than a discrete trial procedure was used and (3) fading was controlled by changes in the subjects' sexual response. Two slide projectors were positioned so that the images overlapped. An electronic sensor detected when the criterion response (70%) was reached. This started a stepping motor which controlled lamp illumination, dimming one lamp and brightening the other. If the erectile response remained above criterion, the stepper continued to operate, bringing the image to full illumination. If it fell below criterion, the fading reversed.

Subjects were asked to pick a child slide that they found highly arousing and an adult slide that they did not find unattractive. The 30-minute session began with a child slide fully illuminated and the adult completely faded out. The sensor was set to detect a 70% erection. Subjects were instructed to fantasize to the slide of the child to produce erection. When fading began and the sexual characteristics of the adult became visible, they were to switch the fantasy to adults, admittedly a difficult requirement. If the subject began to lose his erection and fell below criterion, the adult image faded back to that of the child which would allow him to recapture his erection and fade back to the adult image. A second series of child to adult fading was conducted with different slides. These were randomly selected so that which child would be paired with which adult could not be predicted. Generalization tests were conducted using slides from the fading procedure.

This procedure 'failed to demonstrate the effectiveness of an automated stimulus fading technique. . . . Following child to adult fading, we were unable to show an independent response to adults. After two series of child to adult fading, response to adults presented alone was either lower or insubstantially higher than that shown in the baseline condition' (Laws and Pawlowski, 1974, p. 139).

Automatic Morphing

In recent years, the construction and use of virtual stimuli has appeared. Renaud, Proulx, Rouleau, Bouchard, Bradford, Federoff and Bonin (2006) have demonstrated a procedure where a total change from one stimulus to another can be accomplished through changes in a penile transducer or a vaginal blood flow monitor. Like the fading procedures described previously, the subject first sees an image of an attractive child. In the example presented, when changes occur in the

mercury transducer (Bancroft *et al.*, 1966), the attractive image morphs to an image of an ugly old woman. This is a promising development in that it requires much smaller instrumentation than the procedures reported previously. However, it is also subject to the same problems that befell Laws and Pawlowski (1974). To the author's knowledge, this procedure has not yet been tested clinically.

Virtual Reality and Attention Control Technologies

Renaud (2005) has also reported on a variety of procedures developed in cyberpsychology. He proposes 'a new method for assessing and treating sexual deviance based on the combined use of VR [virtual reality], eye-tracking systems (ETS) and PPG' (Renaud, 2005, p. 1). Renaud argues that the immersive potential of virtual reality significantly increases the external or ecological validity of assessment measures. By 'immersive', he means a 'feeling of presence' (p. 3). This 'feeling of presence' refers to the illusion of being in a real situation of a sexual or potentially sexual nature. . . . The feeling of sexual presence generated by VR can thus increase the validity of sexual interest and preference measures by making the sensorimotor experience in VR more like that outside VR, that is, in the real world' (p. 3). Renaud's subjects enter a virtual environment (VE), a three-sided enclosure in which three projectors back-project still and moving imagery upon the walls. 'The virtual stimuli that we use to elicit sexual attraction and arousal and fully synthetic animated avatars generated with commercial 3D software. These avatars have been designed, developed and validated in order to ensure that they are perceived to represent the required age and sexual properties' (p. 5). The subject is seated inside the VE, wearing a head-mounted display (HMD) which is able to track eye movements. The gaze and dwell time of the eye movements is recorded, providing 'the subjective perspective of the subject being evaluated in virtual immersion' (p. 9). Concurrent PPG measures are also recorded to compare erectile response to the eye-tracking measures.

Conclusions on Innovations

One cannot help but ask why PPG has remained such a stale technology for all these years when potential innovations were obvious but never exploited. The analysis of stimulus content and use of fading procedures never went anywhere. This is unfortunate because they offer a unique approach. The stimulus morphing procedure is similar to the automatic fading procedure but its utility remains to be demonstrated. The use of motion pictures as erotic stimuli has never been repeated. Pornographic videotapes were used for a time in the 1970s but abandoned because it was impossible to know what the subject was responding to. Thus the Freund stimuli remain unique in that the issues of activity and context are controlled. Some innovations are evolutionary such as the pairing of PPG and polygraphy and the multiscreen presentation of visual stimuli along with matching audio scripts. In the author's judgment, virtual reality is going to change everything. The described

26 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

combination of eye tracking and PPG and the use of virtual avatars in a virtual environment represent the new frontier in sexual interest assessment.

COGNITION AND PPG

What is the role of cognition in penile plethysmography? 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.

There are other perspectives on cognition and sexual arousal. For example, Gress (C.L.Z. Gress, personal communication, 29 May 2007) stated:

I think cognition plays a role, but individual differences occur. A person cannot become aroused until they first become aware of and appraise the object of potential interest. Once aware of something, cognition kicks in. I think there are a number of conscious and unconscious . . . regulatory systems that process information and feed that information into a few different feedback loops. The literature suggests one of the loops goes straight to genital arousal, while others cycle through memories and understandings of the current context, which can also feed into genital arousal.

To be sure, cognition has a role in sexual arousal. However, exactly what that role is and exactly how it functions remains to be determined.

SUMMARY ON PPG: A PERSONAL VIEW

In 2003, I published an earlier version of this chapter. It was very negative in tone. I was one of the pioneers in promoting PPG technology in the United States in the 1970s and 1980s. By the turn of the century, I had become quite disillusioned with it, primarily because of its failure to mature and the continuing lack of standardization. I described myself as a former true believer who had become an apostate.

Having written this new version of the material, I am willing to retract some of that negativism. I do not know how it happened but there is finally some movement in this portion of the assessment field. I am especially encouraged by the implementation of the multisite study in the United Kingdom and the detailed procedure manuals that have been developed. This should at least partially solve many of the problems that have been hanging around for decades. I am also encouraged by some of the innovative procedures that have been too long in arriving.

I believe, however, that the future belongs to virtual reality. This is a giant step forward in that it permits us to tailor entire stimulus environments to a particular individual and promises technological evaluation of a host of behaviours, internal as well as external. For many years, these have been the stuff of dreams. Now, at last, they are at hand.

REFERENCES

- Abel, G.G. (1979) Assessment and treatment of child molesters. Grant proposal submitted to the National Institute of Mental Health, No. MH33678-01, Rockville, MD.
- Abel, G.G., Becker, J.V., Card, R.D., Cunningham-Rathner, J., Farrall, W.R., Jensen, S.H., Laws, D.R., Murphy, W.D., Osborn, C.A., Quinsey, V.L. and Wormith, J.S. (1989) *The Stimulus Standardization Study of the Multisite Assessment Group. Paper presented at the First International Conference on the Treatment of Sex Offenders, May 1989, University of Minnesota, Minneapolis, MN.*
- Abel, G.G., Blanchard, E.B., Murphy, W.D., Becker, J.V. and Djenderedjian, A. (1981) Two methods of measuring penile response. *Behavior Therapy*, **12**, 320–28.
- Abel, G.G., Blanchard, E.G., Barlow, D.H. and Mavissakalian, M. (1975) Identifying specific erotic cues in sexual deviation by audiotaped descriptions. *Journal of Applied Behavior Analysis*, **8**, 247–60.
- Association for the Treatment of Sexual Abusers (ATSA) (1997) *Ethical Standards and Principles for the Management of Sexual Abusers. Appendix B: Plethysmograph Examination*, ATSA, Beaverton, OR, pp. 44–51.
- Association for the Treatment of Sexual Abusers (ATSA) (2001) *Practice Standards and Guidelines for Members of the Association for the Treatment of Sexual Abusers. Appendix A: Phallometry*, ATSA, Beaverton, OR, pp. 40–43.
- Bancroft, J.H.J., Jones, H.C. and Pullan, B.P. (1966) A simple transducers for measuring penile erections with comments on its use in the treatment of sexual disorders. *Behaviour Research and Therapy*, **4**, 239–41.
- Barbaree, H.E. and Marshall, W.L. (1988) Deviant sexual arousal, demographic and offense history variables as predictors of reoffense among child molesters and incest offenders. *Behavioral Sciences and the Law*, **6**, 267–80.
- Barker, J.G. and Howell, R.J. (1992) The plethysmograph: a review of the literature. *Bulletin of the American Academy of Psychiatry and Law*, **20**, 13–25.
- Barlow, D.H. (1977) Assessment of sexual behavior, in *Handbook of Behavioral Assessment* (eds A.R. Ciminero, K.S. Calhoun and H.E. Adams), John Wiley & Sons, Ltd, New York, pp. 461–508.
- Barlow, D.H. and Agras, W.S. (1973) Fading in increase heterosexual responsiveness in homosexuals. *Journal of Applied Behavior Analysis*, **6**, 355–66.
- Barlow, D.H., Becker, R., Leitenberg, H. and Agras, W.S. (1970) A mechanical strain gauge for recording penile circumference change. *Journal of Applied Behavior Analysis*, **3**, 72.
- Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579, 113 S. Ct. 2786, 125 L.Ed.2d 469 (1993).
- Day, D.M., Miner, M.H., Sturgeon, V.H. and Murphy, J. (1989) Assessment of sexual arousal by means of physiological and self-report measures, in *Relapse Prevention with Sex Offenders* (eds D.R. Laws), Guilford, New York, pp. 115–23.
- Dorland's Illustrated Medical Dictionary* (1994) 28th edn, W.B. Saunders, Philadelphia, PA.
- Farkas, G.M., Since, L.F. and Evans, I.M. (1979) The effects of distraction, performance demand, stimulus explicitness, and personality on objective and subjective measures of male sexual arousal. *Behaviour Research and Therapy*, **17**, 26–32.
- Freund, K., Sedlacek, F. and Knob, K. (1965) A simple transducer for mechanical plethysmography of the male genital. *Journal of the Experimental Analysis of Behavior*, **8**, 169–70.
- Frye v. US*, 293 F. 1013 (D.C. Cir. 1923).
- Hanson, R.K. and Bussière, M.T. (1998) Predicting relapse: a meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, **66**, 348–62.
- Hanson, R.K. and Morton-Bourgon, K. (2004) Predictors of sexual recidivism: an updated meta-analysis. User Report No. 2004-02. Public Safety and Emergency Preparedness Canada, Ottawa, Ontario.
- HM Prison Service, Offending Behaviour Programmes Unit (2005) *Penile Plethysmograph Procedure (PPG) Documentation*, Home Office, London.
- Howes, R.J. (1995) A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment*, **7**, 9–24.

28 ASSESSMENT OF SEXUAL INTEREST IN SEXUAL OFFENDERS

- Howes, R.J. (2001) Interpretation of low arousal in plethysmographic assessment: an empirical basis. Unpublished manuscript, Stony Mountain Institution, Winnipeg, Manitoba, Canada
- Laws, D.R. (1977) A comparison of the measurement characteristics of two circumferential penile transducers. *Archives of Sexual Behavior*, **6**, 45–51.
- Laws, D.R. (1984) The assessment of dangerous sexual behavior in males. *Medicine and Law*, **3**, 127–40.
- Laws, D.R. (2003) Penile plethysmography: will we ever get it right? in *Sexual Deviance: Issues and Controversies* (eds T. Ward, D.R. Laws and S.M. Hudson), Sage, Thousand Oaks, CA, pp. 82–102.
- Laws, D.R., Gulayets, M.J. and Frenzel, R.R. (1995) Assessment of sex offenders using standardized slide stimuli and procedures: a multisite study. *Sexual Abuse: A Journal of Research and Treatment*, **7**, 155–66.
- Laws, D.R., Hanson, R.K., Osborn, C.A. and Greenbaum, P.E. (2000) Classification of child molesters by plethysmographic assessment of sexual arousal and a self-report measure of sexual preference. *Journal of Interpersonal Violence*, **15**, 1297–312.
- Laws, D.R. and Holmen, M.L. (1978) Sexual response faking by pedophiles. *Criminal Justice and Behavior*, **5**, 343–56.
- Laws, D.R. and Osborn, C.A. (1983) How to build and operate a behavioral laboratory to evaluate and treat sexual deviance, in *The Sexual Aggressor* (eds J.D. Greer and I. Stuart), Van Nostrand Reinhold, New York, pp. 293–335.
- Laws, D.R. and Pawlowski, A.V. (1974) An automated fading procedure to alter sexual responsiveness in pedophiles. *Journal of Homosexuality*, **1**, 149–63.
- Laws, D.R. and Rubin, H.B. (1969) Instructional control of an autonomic sexual response. *Journal of Applied Behavior Analysis*, **12**, 93–99.
- Marshall, W.L. and Fernandez, Y.M. (2000) Phallometric testing with sexual offenders: limits to its value. *Clinical Psychology Review*, **20**, 807–22.
- Marshall, W.L. and Fernandez, Y.M. (2001) Phallometry in forensic practice. *Journal of Forensic Psychology Practice*, **1**, 77–87.
- Marshall, W.L. and Fernandez, Y.M. (2003) Sexual preferences: are they useful in the assessment and treatment of sexual offenders? *Aggression and Violent Behavior*, **8**, 131–43.
- Marshall, W.L., Fernandez, Y.M., Marshall, L.E. and Mann, R.E. (2001) *A proposal to conduct a multi-site standardization study of phallometric assessments*. Unpublished document, Rockwood Psychological Services, Kingston, Ontario, Canada.
- Murphy, W.D. and Barbaree, H.E. (1994) *Assessment of Sex Offenders by Measures of Erectile Response: Psychometric Properties and Decision Making*, Safer Society Press, Brandon, VT.
- Nichols, H.R. and Molinder, L. (1984) *The Multiphasic Sex Inventory Manual*. Available from Nichols and Molinder, 437 Bowes Drive, Tacoma, WA 98466.
- O'Donohue, W. and Letourneau, E. (1992) The psychometric properties of the penile tumescence assessment of child molesters. *Journal of Psychopathology and Behavioral Assessment*, **14**, 123–74.
- Quinsey, V.L. and Chaplin, T.C. (1988) Penile responses of child molesters and normals in descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence*, **3**, 259–74.
- Quinsey, V.L., Chaplin, T.C. and Carrigan, W.F. (1980) Biofeedback and signaled punishment in the modification of inappropriate age preferences. *Behavior Therapy*, **11**, 567–76.
- Quinsey, V.L., Chaplin, T.C. and Varney, G. (1981) A comparison of rapists' and non-sex offenders' sexual preferences for mutually consenting sex, rape, and physical abuse of women. *Behavioral Assessment*, **3**, 127–35.
- Quinsey, V.L., Harris, G.T., Rice, M.E. and Cormier, C.A. (1998) *Violent Offenders: Appraising and Managing Risk*. American Psychological Association, Washington, DC.
- Renaud, P. (2005) The use of virtual reality and attention control technologies in the assessment of deviant sexual preferences. Unpublished manuscript, Université du Québec en Outaouais, Gatineau, Québec, Canada.
- Renaud, P., Proulx, J., Rouleau, J.-L., Bouchard, S., Bradford, J., Federoff, P. and Bonin, M.-P. (2006) *Sexual and Oculomotor Biofeedback Mediated by Sexual Stimuli Presented in Virtual*

PENILE PLETHYSMOGRAPHY: STRENGTHS, LIMITATIONS, INNOVATIONS 29

- Reality*. Paper presented at the annual meeting of the Society for the Scientific Study of Sexuality, November 2006, Las Vegas, NV.
- Rice, M.E., Quinsey, V.L. and Harris, G.T. (1991) Predicting sexual recidivism among treated and untreated child molesters released from a maximum security institution. *Journal of Consulting and Clinical Psychology*, **59**, 381–6.
- Rulo, D.G. (1999) *Can We Identify the Sexual Predator by Use of Penile Plethysmography?* www.forensic-evidence.com/site/Behv_Evid/BeE00005.2.html (accessed 16 June 2001).
- Simon, W.T. and Shouten, P.G.W. (1991) Plethysmography in the assessment and treatment of sexual deviance: an overview. *Archives of Sexual Behavior*, **20** 75–91.
- Smith, S.K. (1998) Evidence of penile plethysmography, psychological profiles, inventories and other 'Not a Pedophile' character and opinion evidence offered on behalf of a defendant in a child sexual abuse case is inadmissible under *Daubert vs. Merrell Dow Pharmaceutical*, <http://www.smith-lawfirm.com/Scientific.Evidence.Brief.html> (accessed 14 June 2001).
- Vogt, W.P. (1993) *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences*. Sage, Newbury Park, CA.
- Wincze, J.P., Wenditti, E., Barlow, D. and Mavissakalian, M. (1980) The effects of a subjective monitoring task in the physiological measure of genital response to erotic stimulation. *Archives of Sexual Behavior*, **9**, 533–45.

