Sexual Abuse: A Journal of Research and Treatment, Vol. 15, No. 4, October 2003 (© 2003)

Circumferential Change Scores in Phallometric Assessment: Normative Data

Richard J. Howes[1,2]

Phallometric testing is a procedure that has enjoyed considerable popularity as an objective component in the assessment of sexual offenders. The value of this procedure may be most notably compromised in the realm of interpretation, and problems in interpretation are particularly acute for those participants where full arousal is not obtained during testing. The calculation of Percent Full Erection (PFE) scores has of necessity involved a speculative component in such cases. Eliminating this speculation through empirical investigation was the purpose of the current research. Circumferential change scores (from flaccidity to full erection) were obtained for 724 respondents at nine North American correctional facilities, allowing for the calculation of descriptive statistics and a determination of the distribution characteristics of these scores. The results provide an empirical basis for calculating PFE scores and interpreting phallometric data in those cases where full arousal is not obtained, and specific confidence levels associated with interpretation are offered. It is suggested that only through a more rigorous application of the principles of science will the procedure of phallometric assessment fulfill its true potential.

KEY WORDS: penile plethysmograph; phallometry; sexual arousal interpretation; sexual offender assessment; phallometric testing.

Many researchers have endorsed the view that the procedure of phallometric assessment has demonstrated both validity and a considerable potential for clinical application (Earls, 1983; Howes, 1995; McAnulty & Adams, 1992; McGrath, 1992; Murphy & Barbaree, 1994; Quinsey, Lalumiere, Rice, & Harris, 1995; Simon & Shouten, 1991; Williams, 1995). Penile tumescence has been described as the most sensitive index of sexual arousal and the most reliable of the physiological

[1]Psychology Department, Stony Mountain Institution, Winnipeg, Manitoba, Canada.
[2]To whom correspondence should be addressed at Psychology Department, Stony Mountain Institution, PO Box 4500, Winnipeg, Manitoba, Canada R3C 3W8; e-mail: howesrj@csc-scc.gc.ca.

1079-0632/03/1000-0365/0

measures (Rosen & Keefe, 1978). Proulx (1989) noted that penile response is the only physiological response specific to sexual arousal in men (i.e., it permits differentiation between sexual arousal and arousal states such as fear and anger). Lalumiere and Harris (1998) commented that "no other measures match the discriminative validity so far achieved by phallometric assessment" (p. 235) and they further noted that few measures match the predictive validity of this technique.

Almost 30 years after his seminal study on the use of phallometric assessment, Freund (1991) ventured that it remains a diagnostic procedure about which there is every reason to be optimistic. Quinsey et al. (1995) reinforced this view by describing phallometric scores as one of the two best theoretically and clinically relevant predictors of sexual reconviction. In their meta-analysis of sexual offender recidivism studies, Hanson and Bussiere (1998) concluded that phallometric measures of arousal to child stimuli are the best predictors of recidivism. Although Maletzky (1991) cautioned that the technique of phallometric assessment is sometimes used indiscriminately in light of inadequate certification standards, an observation later echoed by Laws (1993), he nonetheless noted that numerous clinicians view phallometry as the optimal measure of both normal and aberrant sexual arousal patterns. In its most recent manual on practice standards the Association for the Treatment of Sexual Abusers endorsed the use of phallometric testing and admonished that members should use this procedure to corroborate self-reports regarding sexual arousal patterns (ATSA, 2001).

The apparent worth of phallometric assessment may most notably be diminished in the area of the interpretation of phallometric data. When full arousal to sexual stimuli occurs this does not appear to be a problem of the same magnitude, but difficulty exists in those cases where only low levels of arousal are elicited. No definition of low arousal has been consistently applied in the literature and considerable latitude seems to exist with regard to the meaning of this term, although it has certainly been portrayed to mean arousal which to some degree fails to achieve full erection. In the current study low arousal is operationally defined as any response to phallometric testing which does not reach 20 Percent Full Erection (20 PFE), as the research has established 20 PFE is the most commonly accepted measure for interpreting phallometric data as being clinically significant (see Howes, 1995).

Various explanations might be ventured to account for the phenomenon of low arousal, with anxiety, impression management, and weak stimuli being three likely possibilities. Many clinicians who incorporate phallometric assessment into their work with sexual offenders are familiar with the problem of low arousal, and the research literature is replete with examples (Howes, 1998; Langevin, 1989; Looman, Abracen, Maillet, & DiFazio, 1998; Murphy & Barbaree, 1994). In some cases this problem is substantial, as in the research of Haywood, Grossman, and Cavanaugh (1990) wherein all 51 participants (including 27 control participants) revealed what was described as low arousal. As another example Malcolm, Andrews, and Quinsey (1993) discarded 48% of their sample because of low responding.

As a function of both legal and ethical concerns, considerable energy is currently being expended in examining the issue of what is acceptable in terms of phallometric stimuli. With the distinct possibility that erotic stimuli will become more sanitized and less explicit, perhaps largely or exclusively involving computer-generated images, the issue of low arousal in phallometric assessment may become an even greater problem. Some jurisdictions are relatively immune to legal constraints, most notably in Canada where the use of all forms of otherwise prohibited pornographic materials for scientific or clinical purposes is protected under the law (Government of Canada, 1993). The reality, however, is that increasingly throughout the world clinicians and researchers are likely to be governed by the need to use stimuli which have less power to evoke full arousal, and the need to deal with the problem of interpreting low arousal will become more pronounced.

In an apparent attempt to deal with the issue of low arousal, the practice of calculating *Z*-scores from such data has been advocated by some researchers (Freund, Scher, Rancansky, Campbell, & Heasman, 1986; Quinsey, Chaplin, & Carrigan, 1979). These researchers have offered the suggestion that *Z*-scores are the most sensitive measure in discriminating between sexual offenders and nonoffenders when the scores are used to generate a deviance index. This statistical technique, however, invited some criticism, in part because phallometric responses in individuals are so seldom normally distributed and *Z*-scores are not robust in such a circumstance. Hall (1990) noted that *Z*-score calculations are particularly susceptible to accentuating small differences within subjects, an observation which mirrored a conclusion of Earls, Quinsey, and Castonguay (1987) and which was later endorsed by Shouten and Simon (1992). Hall offered an example where a respondent with tumescence raw scores of 2, 2, 1, 1, 0, 0, 0, and 0 mm would have exactly the same *Z*-scores as a respondent with tumescence raw scores of 20, 20, 10, 10, 0, 0, 0, and 0 mm. He noted that differences between the two response sets were both qualitatively and clinically significant despite the identical *Z*-scores. Langevin (1989) observed that raw scores and *Z*-scores essentially produce the same result, and a perusal of *Sexual Abuse: A Journal of Research and Treatment* reveals that there is simply no recent research which reports phallometric data in terms of raw scores (i.e., millimeters of circumferential change) because of the obvious disadvantages.

Murphy and Barbaree (1994) noted that although *Z*-score transformations account for a greater proportion of the variance this is simply because of the fact that the transformations, by their very nature, eliminate important sources of variance (e.g., between-subjects variability in overall arousal). They also argued that while this method might be statistically powerful for groups of respondents in research settings, a more conservative approach must be adopted in clinical settings in interpreting individual patient data. As well they cautioned about the increased likelihood of committing Type I errors (i.e., false positives) through *Z*-score transformations of data, and it seems apparent that this means of dealing with low arousal is not without difficulty.

The use of PFE scores as the dependent variable has become commonplace (Howes, 1995; Murphy & Barbaree, 1994; Murphy, DiLillo, Haynes, & Sterre, 2001), perhaps in part because this procedure remedies the major criticisms applied to other techniques. PFE calculations are determined on an individual basis but allow for meaningful comparisons with other individuals or groups. The major problem with PFE scores arises in situations of low arousal. In order to avoid using an estimate of the denominator in the calculation, full arousal must be achieved. Although full arousal may be difficult to obtain in some settings, a reliable determination of when full erection has been achieved should not be unduly difficult. Most phallometric technicians have on at least some occasions obtained response tracings which met the criteria described by Furr (1991) where full erection is reflected by clear plateaus and/or consistent peak responses to preferred stimuli (see Howes, 1998). The use of highly explicit materials depicting consenting adult sexual activities (where arousal would be seen by participants as normal and acceptable) clearly assists in a reliable determination of full arousal. Such explicit stimuli have been used or their use advocated by numerous researchers over the years (Abel & Blanchard, 1976; Adams, Motsinger, McAnulty, & Moors, 1992; Howes, 1998; Laws & Osborne, 1983; Wheeler & Rubin, 1987).

When such criteria for full erection are not met, however, definitions of what constitutes full erection have been remarkably inconsistent in the research literature. In spite of axiomatic concerns about the truthfulness of phallometric participants, full erection has been defined in some research according to self-report (Becker, Hunter, Goodwin, Kaplan, & Martinez, 1992a; Kaemingk, Koselka, Becker, & Kaplan, 1995). Using a different criterion, Rea, DeBriere, Butler, and Saunders (1998) reported that for their research sample "Full erection (for each individual) was defined as the largest erection obtained during Study 1" (p. 241), and this value was altered if responding during Study 2 was less or greater for the flaccid and full erection measures respectively.

Specific values estimating circumferential change for PFE calculations have often been used in published studies, and again inconsistency is evident here. Some authors have used 30 mm as an approximation of circumferential change from flaccidity to full erection (Becker, Stern, Kaplan, & Cunningham-Rathner, 1992b; Hunter & Goodwin, 1992), whereas others have reported that full erection "was estimated to be 2.5 cm (25 mm) circumferential change" (Gray, 1995, p. 148). An entirely different assumption was made by Proulx et al. (1997), for they excluded 67 participants from data analysis "because these subjects had not shown penile increases of over 1 mm in diameter, representing approximately 10 to 15% of a full erection, depending on the size of the penis" (p. 10). Proulx et al. were thus venturing that mean circumferential change from flaccidity to full erection was in the range of 6.7 mm (i.e., where 1 mm is 15 PFE) to 10 mm (i.e., where 1 mm is 10 PFE).

Data describing a mean circumferential change score of 24 mm was noted by Furr (1991), whereas Parks Medical Electronics recommended estimating the

change from flaccidity to full erection as 30 to 40% of the flaccid baseline (Parks Medical Electronics, Inc., 1983). As an example of the Parks estimation, an individual with a flaccid penis circumference of 90 mm would be expected to reveal a circumferential change score of from 27 mm to 36 mm. Clearly the entire area of determining circumferential change to allow for PFE calculations in the absence of full erection is characterized by a substantial degree of variability. The number of studies where PFE scores are reported even in the absence of full arousal, often with idiosyncratic calculation methods, does not reflect well on the state of the literature. This variability may actually compromise the acceptance of phallometry as a legitimate assessment technique.

The actual calculation of PFE scores is quite straightforward, with the increase in arousal to a stimulus being the numerator and the circumferential change between flaccidity and full erection being the denominator in the equation. Such calculations are of course specific to individuals, as befits naturally existing differences in individual penis size and circumferential change during sexual arousal. When a participant does not achieve a full erection, however, any confidence in the accuracy of PFE calculations is diminished by uncertainty about what the denominator should be. This is the crux of the problem, since without full erection the PFE calculations are only speculative. Minimizing this speculative component is the intent of this current research, for the determination of circumferential change scores clearly lends itself to empirical investigation. With a large enough sample of participants who achieve full erection the calculation of the mean and standard deviation should allow for the establishment of confidence levels for PFE scoring and interpretation.

METHOD

Fourteen North American correctional facilities and treatment centers that were known to process large numbers of sexual offenders were contacted to seek phallometric testing data for this study. Nine of these centers readily agreed to participate and ultimately provided data, two agreed to participate but never provided data, two of the fourteen centers did not return phone calls, and one center reported that because of increasing liability insurance costs they had abandoned phallometric testing in their facility.

The specific data sought were the circumferential change scores (in millimeters) from flaccidity to full erection of those individuals for whom it could be stated with every reasonable confidence that full erection had been obtained during plethysmographic testing. Although some data were reported from cases where visual observation confirmed the presence of full erection, in most cases satisfaction of the qualification “every reasonable confidence” had to be based on response tracings which clearly revealed consistent plateaus to repeated presentations of preferred stimuli. Such a definition of full erection is consistent with

the criteria described by Furr (1991). Participants' self-reports were not used in the determination of the presence of full erection, although in many cases such self-reports were consistent with the response tracings.

No data were sought with regard to the nature of the phallometric stimuli that evoked arousal in the participants. Whether or not the reported circumferential change scores were in response to consenting adult heterosexual stimuli or pedophilic stimuli was simply not germane to the current research issue.

Data from 699 participants where full erection was obtained were provided by nine contributing facilities. Data from one of these facilities ($N = 84$) were eliminated from the final analysis because initial analysis revealed the idiosyncratic nature of these data (see subsequent note). An additional 109 cases were available from the author's institution, yielding a total sample size of 724 male participants from nine sites.

RESULTS

The mean circumferential change score for the sample of 724 participants was 32.57 mm with a mode of 30 mm, a median of 31 mm, and a standard deviation of 8.76 mm. As indicated in Fig. 1 the scores approximated a normal distribution, although predictably there was a moderate positive skewness (i.e., 0.5698) to the distribution. As well the distribution was mildly leptokurtic (i.e., kurtosis = 0.2856). Consistent with the characteristics of a normal distribution 95% of the actual scores fell at or below 49 mm and 99% of the actual scores fell at or below 56 mm.

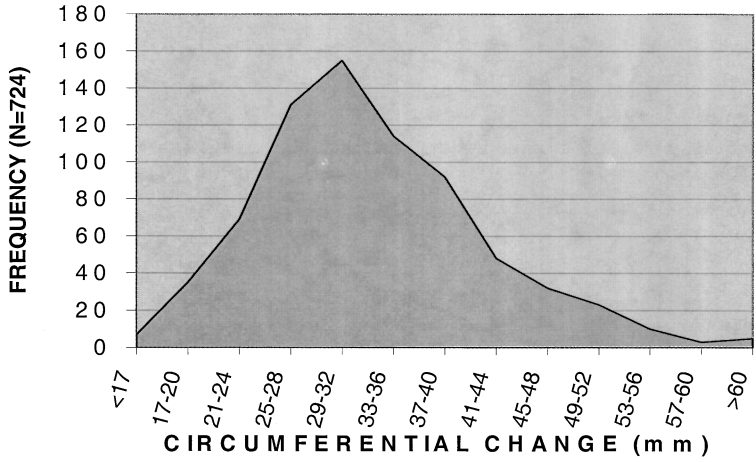

Fig. 1. Frequency distribution of circumferential change scores.

Table I. Descriptive Statistics of Groups

Group	*N*	Mean	Standard deviation	Skewness	Kurtosis
A	78	34.38	7.34	0.3710	0.6803
B	86	30.22	10.65	0.9324	0.5832
C	79	35.65	9.72	0.3185	0.1499
D	63	30.28	7.72	0.7566	0.7875
E	109	32.03	6.97	0.7577	0.2208
F	189	31.63	9.13	0.7294	0.4513
G	53	36.88	8.04	0.1152	0.1414
H[a]	67	32.17	8.92	0.5279	0.2967
X[b]	84	52.98	16.45	−0.2078	−0.5688

[a] Combines small samples (<50).
[b] Eliminated from final analysis.

It should be acknowledged that statistically significant differences were observed between the means of some samples, although this may have been more a function of large sample sizes accentuating small differences and experimentwise error rate than a reflection of any real differences of practical significance. Of the 28 possible comparisons between individual group means, 11 of the differences were significant at the .05 probability level. The actual range of sample means in the eight data sets included for analysis (i.e., two groups of less than 50 respondents were combined) was from 30.22 mm to 36.88 mm with standard deviations ranging from 6.97 to 10.65 mm (see Table I). The possibility that the observed sample differences might be explained by differences in the training of phallometric technicians must certainly be considered.

An initial analysis of the one sample ultimately eliminated from this study revealed extreme differences from all other samples, differences which went far beyond simply being statistically significant. The mean of this sample of 84 participants was determined to be 52.98 mm with a standard deviation of 16.45 mm in what was the only platykurtic distribution (i.e., kurtosis $= -0.5688$) and the only negatively skewed distribution (i.e., skewness $= -0.2078$). A comparison of this sample with the sample closest in value (i.e., 36.88 mm) revealed a t statistic at approximately four times the level necessary for significance at the .001 probability level ($t = 7.198$, $p < .001$). These data were excluded from further analysis to avoid contaminating the remaining data.

DISCUSSION

Given the large sample size and the diverse sources of these data one can have reasonable confidence in generalizing the results to the larger male sexual offender population. Such confidence is promoted when we consider that what is perhaps the most commonly used psychometric instrument (i.e., the MMPI-2) is based on a normative sample of 1138 men (and 1462 women) from seven states (see Pope,

Butcher, & Seelen, 1993). The current sample of 724 men from nine provinces and states would seem to fare reasonably well in comparison to this well-established instrument.

The current research allows for the conclusion that circumferential change scores from flaccidity to full erection for male sexual offenders have a mean of 32.6 mm and a standard deviation of 8.8 mm and that these scores are normally distributed. Inasmuch as there is an acknowledged normal distribution in most other naturally occurring physical characteristics which have a ratio level of measurement (e.g., height, weight, foot size) the phenomenon measured here is entirely predictable. The range and variability of scores obtained is also consistent with previous research which revealed that half of the 48 phallometric practitioners who responded to a survey indicated that they had encountered circumferential change scores of greater than 50 mm in at least one case (see Howes, 1995). Likewise 80% of these practitioners indicated that they had encountered circumferential change scores to full erection of less than 20 mm.

Knowing the mean and standard deviation in a normal distribution allows us to arrive at some descriptive and predictive conclusions. Ninety-five percent of circumferential change scores can be expected to fall at or below 47 mm (i.e., 1.64 Z-scores above the mean), and 99% can be expected to fall at or below 53 mm (i.e., 2.33 Z-scores above the mean). Confidence levels can thus be established with regard to phallometric data in the absence of full erection. If we can assume that the most commonly accepted figure of 20 PFE reflects clinical significance in phallometric testing, as noted earlier, then even without a full erection we can conclude that 19 times out of 20 a circumferential change score of 9.4 mm (i.e., 20% of the 95th percentile score in the distribution) is significant and therefore interpretable. There will certainly be cases where circumferential change of a lower magnitude is significant, but in the absence of full erection there is an unacceptable risk of committing a Type 1 error if at least 9.4 mm of change has not been demonstrated. In a similar manner, a demonstrated circumferential change score of 10.6 mm (even in the absence of full erection) can be regarded as significant and interpretable at the .01 confidence level, for this value is 20% of the 99th percentile score in the distribution. Such a level of confidence would certainly satisfy the "beyond a reasonable doubt" admonition in the criminal justice system, a consideration of no small value given the number of phallometric practitioners who have been qualified as expert witnesses in court.

Wilson, Abel, Coyne, and Rouleau (1992) observed that "millimeters change gives no indication of any one individual's relative degree of sexual arousal" (p. 164), and this has thusfar been a valid statement whenever full erection has not been obtained. The current data, however, allow us to interpret millimeters of circumferential change in those cases where indications of full erection are uncertain or absent. Even if an individual never achieves arousal beyond 11 mm of change we can with some confidence ($p < .01$) conclude that this arousal is

clinically significant and proceed to make appropriate interpretations. The level of confidence and thus the criterion circumferential change score might be lowered in cases where the consequences of Type 1 error are not dramatic, with research studies being an obvious example. In those situations where false positive interpretations may have dire consequences such as extended incarceration, however, then compromising conservative confidence levels invites criticism. As noted earlier the figure of 30 mm has frequently been used as the denominator in PFE calculations in the absence of full erection, although the current data reveal that this figure is actually at the 48th percentile of the distribution. It would seem that the error rate associated with using anything less than 47 mm as the denominator (when full arousal is not demonstrated) is simply unacceptable by conventional scientific standards.

There is not much disagreement that phallometric assessment is a legitimate measure of sexual arousal, but inconsistencies in scoring and interpretation may compromise the worth of this procedure. Data that appear to have little statistical or practical significance have been interpreted as meaningful in the published research, and one can only speculate about how often such interpretations have been applied in clinical and court settings. As well, variability in criterion levels for the significance of data should clearly not exist if the procedure of phallometric assessment is to gain credibility. The current data not only offer an empirical basis for determining when sexual arousal reaches significance at specific confidence levels, they also allow for the interpretation of phallometric data when full arousal is not obtained.

Whether or not phallometric assessment will become simply an interesting but irrelevant electronic curiosity is yet to be determined, but such a fate can best be avoided through the proper application of the principles of science. The potential for this procedure to assist in the assessment and treatment of sexual offenders exists, but only with rigorous and objective standards can its value be confirmed.

ACKNOWLEDGMENTS

The author acknowledges the assistance of the individuals from many sites who provided additional data for this research. Andy Haynes of the Regional Psychiatric Centre, Saskatchewan, Tom Murray of Bowden Institution, Alberta, Michael Black of the Regional Reception Centre, Nova Scotia, Rae King of Dorchester Penitentiary, New Brunswick, Todd Smith of the Forensic Behavioral Management Clinic, Manitoba, Dr Brian Nadeau, Regional Psychiatric Centre, BC, and Greg Maillet, Regional Treatment Centre, Ontario, all contributed in this manner. As well, John Davis, Clinical Director of RSA Inc. in Lakewood Colorado, and Rick Minnich of Minnich Polygraph Service, Washington, provided something of an international aspect to the research sample. Gratitude is also expressed to

the Correctional Service of Canada for supporting this research endeavor and to Dr Kent Somers in particular for authorizing time to complete it. The opinions expressed in this study are those of the author and do not necessarily reflect the opinions or policies of the Correctional Service of Canada.

REFERENCES

Abel, G. G., & Blanchard, E. B. (1976). The measurement and generation of sexual arousal in male sexual deviates. *Progressive Behavior Modification, 2,* 99–136.

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.

Association for the Treatment of Sexual Abusers (ATSA) (2001). *Practice standards and guidelines for Members of the Association for the Treatment of Sexual Abusers*. Beaverton, OR: Author.

Becker, J. V., Hunter, J. A., Goodwin, D., Kaplan, M. S., & Martinez, D. (1992a). Test–retest reliability of videotaped phallometric stimuli with adolescent sexual offenders. *Annals of Sex Research, 5,* 45–51.

Becker, J. V., Stein, R. M., Kaplan, M. S., & Cunningham-Rathner, J. (1992b). Erection response characteristics of adolescent sex offenders. *Annals of Sex Research, 5,* 81–86.

Earls, C. M. (1983). Some issues in the assessment of sexual deviance. *International Journal of Law and Psychiatry, 6,* 431–441.

Earls, C. M., Quinsey, V. L., & Castonguay, L. G. (1987). A comparison of three methods of scoring penile circumference changes. *Archives of Sexual Behavior, 16,* 493–500.

Freund, K. (1991). Reflections on the development of the phallometric method of assessing sexual preferences. *Annals of Sex Research, 4,* 221–228.

Freund, K., Scher, H., Rancansky, I., Campbell, K., & Heasman, G. (1986). Males disposed to commit rape. *Archives of Sexual Behavior, 15,* 23–35.

Furr, K. D. (1991). Penis size and magnitude of erectile change as spurious factors in estimating sexual arousal. *Annals of Sex Research, 4,* 265–279.

Government of Canada. (1993). *Criminal code of Canada.* Agincourt, ON: Carswell.

Gray, S. R. (1995). A comparison of verbal satiation and minimal arousal conditioning to reduce deviant arousal in the laboratory. *Sexual Abuse: A Journal of Research and Treatment, 7,* 143–153.

Hall, G. C. (1990). Validity of pedophilic sexual arousal in a sexual offender population: A reply to Quinsey and Laws. *Journal of Consulting and Clinical Psychology, 58,* 889–891.

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.

Haywood, T. W., Grossman, L. S., & Cavanaugh, J. L. (1990). Subjective versus objective measurements of deviant sexual arousal in clinical evaluations of alleged child molesters. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 269–275.

Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment, 7,* 9–24.

Howes, R. J. (1998). Plethysmographic assessment of nonsexual offenders: A comparison with rapists. *Sexual Abuse: A Journal of Research and Treatment, 10,* 183–194.

Hunter, J. A., & Goodwin, D. W. (1992). The clinical utility of satiation therapy with juvenile sexual offenders: Variations and efficacy. *Annals of Sex Research, 5,* 71–80.

Kaemingk, K. L., Koselka, M., Becker, J., & Kaplan, M. S. (1995). Age and adolescent sexual offender arousal. *Sexual Abuse: A Journal of Research and Treatment, 7,* 249–257.

Lalumiere, M. L., & Harris, G. T. (1998). Common questions regarding the use of phallometric testing with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10,* 227–237.

Langevin, R. (1989). *Sexual preference testing.* Toronto: Juniper Press.

Laws, D. R. (1993). The President's column: Certification of plethysmographic operators. ATSA Newsletter, *5,* 2–4.

Laws, D. R., & Osborne, 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*, New York: Van Nostrand.

Looman, J., Abracen, J., Maillet, G., & DiFazio, R. (1998). Phallometric nonresponding in sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10,* 325–336.

Malcolm, P. B., Andrews, D. A., & Quinsey, V. L. (1993). Discrimination and predictive validity of phallometrically measured age and gender preferences. *Journal of Interpersonal Violence, 8,* 486–501.

Maletzky, B. M. (1991). *Treating the sexual offender*. Newbury Park, CA: Sage.

McAnulty, R. D., & Adams, H. E. (1992). Validity and ethics of penile circumference measures of sexual arousal: A reply to McConaghy. *Archives of Sexual Behavior, 21,* 177–186.

McGrath, R. J. (1991). Sex-offender risk assessment and disposition planning: A review of empirical and clinical findings. *International Journal of Offender Therapy and Comparative Criminology, 35,* 328–350.

Murphy, W. D., & Barbaree, H. E. (1994). *Assessments of sex offenders by measures of erectile response: Psychometric properties and decision making.* Brandon, VT: Safer Society Press.

Murphy, W., DiLillo, D., Haynes, M., & Steere, E. (2001). An exploration of factors related to deviant sexual arousal among juvenile sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 13,* 91–103.

Parks Medical Electronics, Inc. (1983). *Operating manual: Model 240 strain gauge plethysmograph*. Beaverton, OR: Author.

Pope, H., Butcher, J., & Seelen, J. (1993). *The MMPI, MMPI-2, and MMPI-A in court*. Washington, DC: American Psychological Association.

Proulx, J. (1989). Sexual preference assessment of sexual aggressors. *International Journal of Law and Psychiatry, 6,* 431–441.

Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J., & Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual aggressors. *Sexual Abuse: A Journal of Research and Treatment, 9,* 7–27.

Quinsey, V. L., Chaplin, T. C., & Carrigan, W. F. (1979). Sexual preferences among incestuous and nonincestuous child molesters. *Behavior Therapy, 10,* 562–565.

Quinsey, V. L., Lalumiere, M. L., Rice, M. E., & Harris, G. T. (1995). Predicting sexual offences. In J. C. Campbell (Ed.), *Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers.* Thousand Oaks, CA: Sage.

Rea, J. A., DeBriere, T., Butler, K., & Saunders, K. J. (1998) An analysis of four sexual offenders' arousal in the natural environment through the use of a portable penile plethysmograph. *Sexual Abuse: A Journal of Research and Treatment, 10,* 239–255.

Rosen, R. C., & Keefe, F. J. (1978). The measurement of human penile tumescence. *Psychophysiology, 15,* 366–376.

Shouten, P. G. W., & Simon, W. T. (1992). Validity of phallometric measures: Comments on the Quinsey, Laws, and Hall debate. *Journal of Consulting and Clinical Psychology, 60,* 812–814.

Simon, W. T., & Shouten, P. G. W. (1991). Plethysmography in the assessment and treatment of sexual deviance: An overview. *Archives of Sexual Behavior, 20,* 75–91.

Wheeler, D., & Rubin, H. B. (1987). A comparison of volumetric and circumferential measures of penile erection. *Archives of Sexual Behavior, 16,* 289–299.

Williams, S. M. (1995). Sex offender assessment guidelines. In T. Leis, L. Motiuk, & J. Oglof (Eds.), *Forensic psychology: Policy and practice in corrections*. Ottawa, ON: Correctional Service of Canada.

Wilson, K., Abel, G., Coyne, B., & Rouleau, J. (1992). Sex guilt and paraphilic behavior. *Annals of Sex Research, 5,* 161–170.