Psychophysiological Signs of Faking in the Phallometric Test

Robin J. Wilson[1]

Breathing and finger pulse rates were investigated as potential indicators of faking in a phallometric test procedure. Forty-nine heterosexual male university students with no admitted or documented history of sexual deviance were divided into two groups and administered a standard phallometric test for age and gender preference according to the following procedures: (1) instructed faking of a preference for female children (N = 23) and (2) a no-instruction control (N = 26). Results confirmed earlier research showing that the phallometric test was easily falsified. With respect to breathing and finger pulse, statistically significant positive correlations between penile volume and finger pulse rate were observed in the instructed faking group during conditions of faking. No significant correlations among the three measures (penile tumescence, finger pulse rate, and breathing rate) were found in the no-instruction control group. Subjects in the faking group were also assessed with respect to perceived success in response enhancement or suppression. Results are summarized with respect to clinical relevance.

KEY WORDS: assessment; faking; false positives; penile plethysmograph.

INTRODUCTION

The objective diagnosis of paraphilic tendencies in sexual offenders most often requires the use of penile plethysmography or, as it has been known regarding volumetric measurement, the phallometric test (Freund, 1957). The phallometric procedure involves measuring changes in penile dimensions during presentations of audiovisual stimuli. This may be accom-

[1]Relapse Prevention Maintenance Program, Correctional Service of Canada, Central Ontario District, 330 Keele Street, Main Floor, Toronto, Ontario M6P 2K7, Canada.

plished by measuring either penile circumference (Bancroft, Jones, & Pullan, 1966; Barlow, Becker, Leitenberg, & Agras, 1970) or penile volume (Freund, Sedlacek, & Knob, 1965). Presently, the great majority of phallometric laboratories employs a circumferential measure consisting of a mercury-in-rubber strain gauge.

Since its inception, many researchers have attempted to establish the strengths and limitations of the phallometric test as a method of diagnosis for sexual offenders. While the test has not been universally standardized, particular groups of clinicians have published results outlining the psychometric properties of their own standard procedures (see Freund & Blanchard, 1989; Freund & Watson, 1991; Harris, Rice, Quinsey, Chaplin, & Earls, 1992). While results have varied, it would seem that phallometric procedures for diagnosing age and gender preferences are both reliable and valid, particularly for outpatient admitters, but also for nonadmitters (Freund, Chan, & Coulthard, 1979).

For example, Freund and associates (Freund & Blanchard, 1989; Freund & Watson, 1991) have reported that the phallometric test of gender and age preferences is both reliable and valid. The outcomes of these two controlled studies indicated that the test was 95% specific, while generating sensitivities of greater than 75% for heterosexual pedophiles and greater than 85% for homosexual pedophiles. It proved more difficult to separate opportunistic offenders from pedophiles in the heterosexual offenders than in homosexual offenders. Nonetheless, the test appeared to be valuable as a diagnostic tool with sexual offenders against children.

Results of studies focusing on activity preference disorders, such as rape proneness, paraphilic sadism, and the variants of courtship disorder (Freund & Watson, 1990), however, have been less consistent, suggesting a need for further investigation, such as the meta-analytic approach taken by Lalumiére and Quinsey (1993, 1994).

As the literature expounding the usefulness of phallometry in the diagnosis of paraphilias has grown, so too has the literature on its limitations. One of these is its susceptibility to interference by the test subject (faking). As early as 1963, the ability of phallometric subjects to fake responses was noted (Freund, 1963). A number of researchers have subsequently added empirical data to support the claim that the test is sometimes easily faked (Henson & Rubin, 1971; Laws & Holmen, 1978; Laws & Rubin, 1969; Malcolm, Davidson, & Marshall, 1985; Rubin & Henson, 1975). Self-reports by tested individuals have indicated that, aside from physical means (e.g., perineal muscle contractions or "pumping"), manipulation of the test can best be achieved by ignoring certain parts of the stimuli or by replacing them with imagery that would serve to augment or suppress responding. Self-reports have confirmed that many test subjects attempt to increase

arousal to desirable stimuli, such as female adults. Indeed, Freund, Watson, and Rienzo (1988) noted that perineal muscle contractions are found almost exclusively during presentation of stimuli depicting women.

Freund and associates (1988) delineated three signs of faking for age and gender preference: (1) pumping, (2) preferential response to sexually neutral stimuli, and (3) lack of discrimination between male and female stimuli, where at least one of the two stimulus categories was of adult persons. They also reported that it was easier for their subjects to fake an age preference than it was to fake a spurious gender preference and that success in faking was related to subjects' familiarity with the test procedure as well as the particular test procedure employed. These findings confirmed earlier results (Laws & Holmen, 1988; Rubin & Henson, 1975) indicating that both response suppression and arousal production were possible and that it would be easier for a nonpedophilic heterosexual male to feign a preference for female than for male children.

The battle against faking in phallometric testing has been fought on a number of fronts, but continues to present problems for clinicians. Attempts have been made to increase stimulus strength by using more explicit stimulus materials. Freund (1965) reported methods to increase the impact of stimuli, such as giving an injection of testosterone the evening before testing and serving alcohol prior to the test. The subjects' concentration to test stimuli has also been targeted in a number of attention-monitoring procedures; for example, subjects were required to follow a pattern of flickering lights presented on the same screen as the test stimuli. However, most attention tasks seem only to compete with the test process (Adams, Motsinger, McAnulty, & Moore, 1992). Quinsey and Chaplin (1988) have described a "semantic tracking task" in which subjects pressed buttons whenever they encountered certain critical audiotaped components, such as sexual interaction and violence. However, attempts to replicate this study have not been successful (McAnulty & Adams, 1991).

Another attempt to thwart faking involved the use of other psychophysiological measures, such as those used in polygraphy (Kaine, Crim, & Mersereau, 1988). However, several difficulties have been identified in using such measures in conjunction with phallometry. First, some distinction should be made between general and specific arousal. There has been little success in efforts reliably to diagnose paraphilic preferences in men using any method other than penile plethysmography (Geer, 1975; Masters, Johnson, & Kolodny, 1994). However, while *differential* diagnosis might not be possible using other psychophysiological measures, such measures might provide data about attempts to alter responding.

Some researchers (Oswald & Cleary, 1986; Rosen, Shapiro, & Schwartz, 1975) have used heart and respiration rates in their work with phallometry

and have found that, under conditions of tension, heart rate tends to increase while breathing becomes erratic. A faked response could be interpreted as one in which the subject experienced some degree of tension. We sought to replicate this finding in a study of instructed faking with university students utilizing pulse and breathing rates.

METHOD

Subjects

Two groups of male university students were recruited from a student job placement center and by advertisements posted in common student areas. Both undergraduates and graduates responded to the postings, and virtually all respondents were Caucasians within a range of 20 to 30 years old.

Subjects in both groups were screened on initial telephone contact about their gender preference and were subsequently screened by questionnaire about involvement in sexually deviant behaviors. Only heterosexual men with no admitted history of sexually deviant acts were selected. Group 1 (instructed fakers) consisted of 23 students, while Group 2 (control group) consisted of 26 students. All subjects were paid $20 for their participation and informed consent was obtained.

Procedure

Phallometric Testing and Apparatus. Participants in both subject groups were administered Session 2 of a standard phallometric test for age and gender preference (for a more detailed description see Freund & Watson, 1991). Stimulus categories included male and female children (mean age, approximately 8), male and female adults, and sexually neutral stimuli. The test consisted of 5 blocks of 5 trials, for a total of 25 trials. Each block included one trial of each stimulus category in fixed but pseudo-random order and each lasted 54 sec. Stimulus materials were composed of slides depicting three views (full-figure front, full-figure rear, and genital region) of nude models and accompanying sexually explicit audiotaped narratives appropriate to the age and gender of the slides.

The Freund volumetric penile plethysmograph was used (Freund *et al.*, 1965). In addition to the volumetric phallometer, each subject was fitted with an IBS Corporation Biomedical Instruments Inc. Pulse Rate Monitor (Model CT-1600). The breathing rate apparatus was constructed from a Fenwall glass-coated bead thermistor (Part No. 112-103DJ-B01) attached

to headphones in the same manner as a telephone operator's microphone but situated near the subject's nostril. Subjects were instructed not to touch the thermistor and to breathe normally.

Measurement and Data Conversion. In this study, three types of physiological data were sampled: (1) penile blood volume, the dependent measure of erotic arousal; (2) finger pulse rate, assumed to be a dependent measure of faking under conditions of stress; and (3) breaths per trial, also assumed to be a dependent measure of faking under conditions of stress.

Before presentation of any test stimuli, baseline penile volume was established, a process which lasted about 5 min. Baseline was determined as that level of penile volume from which the subject did not deviate appreciably (±0.50 ml) while discussing sexually neutral topics and while not being presented with sexual stimuli.

For each trial, changes in penile volume (ml) were measured in two ways: (1) a D-type score—the largest deviation from initial value (or DMAX); and (2) an A-type score—the area under the plotted test curve. These scores were then converted into standard scores derived from each subject's own raw D- and A-type scores. These individual standardized scores were then combined according to the formula $(zD + zA)/2$. Area under the curve and maximum deviation should be highly correlated, and any discrepancy between these two scores would likely be representative of an irregularity, as, for instance, body movement (Freund & Watson, 1991). A combined D-A score lowers the impact of such irregularities in the evaluation of test results.

Subjects were required to produce a minimum level of arousal in order for the test procedure to be considered valid. The *output index* (OI) routinely used in our labs is 1.0 ml and is calculated by averaging a subject's three largest DMAX scores (not including responses to sexually neutral stimuli). In any test session with a valid OI, an individual's gender and age preferences may be calculated by comparing his responses, males vs. females and children vs. adults. This procedure has been validated in two controlled studies (Freund & Blanchard, 1989; Freund & Watson, 1991). It is not necessary to gauge the magnitude of response in comparison to other individuals—it is the ratio between stimulus categories that is most crucial. In this study, the OIs generated by members of the instructed faking group did not differ substantially in magnitude from those generated by the control group.

Finger pulse was continuously measured during the procedure and the data were reduced in the same fashion as the phallometric data. Measurement of breathing rate resulted in only one score, number of breaths per trial. This score was then averaged within each stimulus category.

Instructed Faking Group. Subjects in the instructed faking procedure were told that they were part of an investigation aimed at establishing the

validity of a psychophysiological test of pedophilia. They were told further that we were concerned about the susceptibility of this test to faking and that they were to attempt to fake a preference for female children according to our instructions. Participants were asked to elevate their responses to female children while suppressing responses to all other stimulus categories. They were also cautioned that their efforts should be limited to cognitive means (e.g., imagery, fantasy, etc.), as any attempts using physical methods would be readily seen on plots of response curves.

Following the phallometric procedure, all instructed fakers were asked whether they believed their attempts at faking a preference for female children had been successful. They were also asked what means they had employed in this regard.

No-Faking Control Group. Subjects in the control group were given the same instructions except that they were not given any instructions with respect to faking. Phallometric, finger pulse, and breathing rate devices were all employed in the same fashion as with the faking group.

RESULTS

All statistical procedures were accomplished by computer program (SPSS, 1991). Statistical significance in all cases was taken as $p < .05$. Due to technical difficulties, three subjects in the faking group had unusable phallometric and finger pulse data, while the same was true of three individuals in the control group. One additional subject in the control group had unusable finger pulse data and four subjects had unusable breathing rate data. Four subjects in the control group also failed to demonstrate enough arousal for the phallometric test to be considered valid. Only invalidation of phallometric data resulted in a subject being entirely excluded from the study.

The instructed fakers were significantly able to increase their responses to female children and suppress responding to female adults compared to their control counterparts $[t(37) = 3.10, p < .01,$ and $t(37) = -2.45, p < .05$, respectively, for raw DMAX scores]. Levels of arousal (mean and standardized maximum deviation from baseline and area under the curve) for these two stimulus categories were nearly inverse between the two subject groups. The differences in responding to female children and female adults within each subject group were also statistically significant $[t(19) = 2.93, p < .01,$ and $t(18) = -3.98, p < .005$, respectively, for fakers and non-fakers]. The groups did not differ in responses to any of the other stimulus categories, either between or within the subject groups. These results are summarized in Table I.

Table I. Mean Arousal Scores (SD)—DMAX and Area Under Curve

	Stimulus category				
	FC*	FA*	MC	MA	N
DMAX					
Fakers	9.56 (8.22)	4.11 (4.75)	-.13 (1.42)	-.49 (1.26)	-.10 (1.80)
Controls	3.35 (4.63)	8.99 (7.72)	.51 (1.68)	.11 (1.63)	-.67 (1.14)
Z-DMAX					
Fakers	1.27 (.57)	.28 (.62)	-.44 (.33)	-.51 (.29)	-.60 (.27)
Controls	.11 (.50)	1.32 (.47)	-.40 (.17)	-.40 (.32)	-.62 (.38)
Area (× .01)					
Fakers	2.24 (2.22)	1.02 (1.13)	-.14 (.32)	-.21 (.36)	-.02 (.54)
Controls	.62 (1.06)	2.29 (2.22)	-.02 (.33)	-.06 (.29)	-.17 (.29)
Z-area					
Fakers	1.24 (.58)	.35 (.60)	-.51 (.32)	-.55 (.27)	-.53 (.24)
Controls	.09 (.49)	1.30 (.55)	-.41 (.19)	-.41 (.37)	-.56 (.40)

Note. DMAX, maximum deviation from baseline (ml); Z-DMAX, standard score DMAX; Area, area under the curve (scores multiplied by .01); Z-area, standard score AREA; FC, female children; FA, female adults; MC, male children; MA, male adults; N, neutral stimuli.
*The groups differed significantly on scores for FC ($p < .01$) and for FA ($p < .05$) on each of the four measures of arousal.

Instructed Faking Group

Seventeen of the 20 subjects with valid phallometric outcomes in the faking group were successful in faking a preference for female children or were at least successful in raising their arousal to girls to a level equivalent to their responses to women.

Table II reports phallometric, finger pulse rate, and breathing responses in both groups. The instructed faking group showed a noticeably higher response to female children than did their noninstructed counterparts as measured by both the phallometric and the finger pulse devices. Mean breathing rate did not distinguish between the groups.

Correlation results for the groups are summarized in Table III. Statistically significant positive correlations with responses to female children were noted between finger pulse and each of the other two measures. Inspection of trial plots for phallometric and finger pulse measures within any successful faker revealed similarities. Figure 1 demonstrates, for an individual subject, the typical concordance between faked phallometric and finger pulse responses to female child stimuli. A significant negative cor-

Table II. Mean zPHM, zFPR, and BR—Fakers and Controls

Stimulus	zPHM	zFPR	BR
FC			
Fakers	1.269	.612	13.650
Controls	.107	-.156	13.076
FA			
Fakers	.281	-.028	13.063
Controls	1.324	.306	13.323
MC			
Fakers	-.438	-.262	12.811
Controls	-.402	-.122	12.748
MA			
Fakers	-.512	-.219	12.919
Controls	-.404	-.142	13.128
N			
Fakers	-.601	-.104	12.448
Controls	-.625	.115	13.207

Note. zPHM, mean standardized phallometric response; zFPR, mean standardized finger pulse response; BR, mean breathing rate (per trial—54 sec); FC, female children; FA, female adults; MC, male children; MA, male adults; N, sexually neutral.

Table III. Correlation Coefficients—Fakers and Controls

Stimulus	PHM/FPR	PHM/BR	FPR/BR
MC			
Fakers	-.278	.007	-.331
Controls	-.065	.163	.211
FC			
Fakers	.486*	.331	.598*
Controls	.126	-.469	.040
FA			
Fakers	.088	-.516*	.148
Controls	-.328	-.051	.144
MA			
Fakers	.108	.193	.149
Controls	-.225	-.219	-.321
N			
Fakers	.218	.392	.195
Controls	.081	.150	-.544*

Note. PHM, mean phallometric response; FPR, mean finger pulse response; BR, mean breathing rate; FC, female children; FA, female adults; MC, male children; MA, male adults; N, sexually neutral.
*$p < .05$.
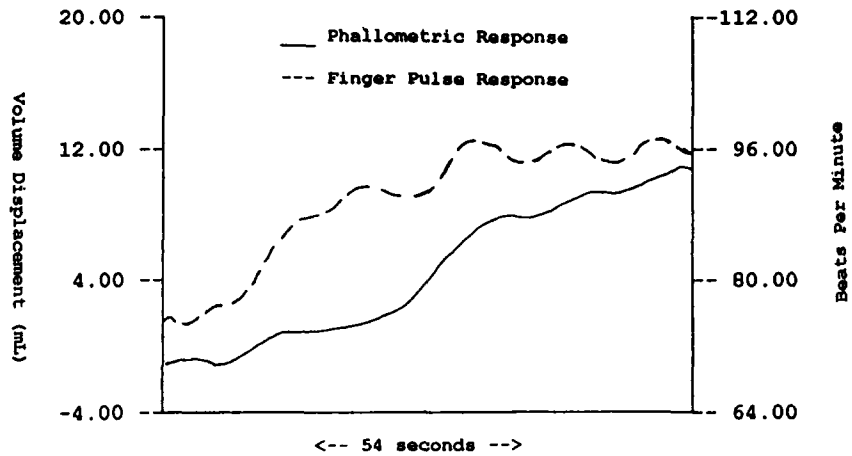
Fig. 1. Phallometric and finger pulse responses in a faking subject.

relation was noted between the phallometric and the breathing measures for response to female adults within this group.

Subjects in the faking group reported a number of techniques for enhancing or suppressing responses. The most common means of enhancing response to female children was to ignore the model presented and to fantasize about interactions with an adult. The most common suppression tactic was to think about something unpleasant. Of the 17 subjects who were able to feign a preference for female children, or who were at least able to raise their responses to female children to a level equivalent to their responses to female adults, 13 believed they were successful in faking, while only 1 thought he was not. Of three subjects not successful at faking, two thought they had been successful while the other realized he had not.

No-Instruction Control Group

Of the 19 students in the control group with valid phallometric outcomes, 16 generated a gynephilic phallometric profile. This is consistent with results published previously (Freund & Watson, 1991) in which approximately 20% of university student controls tested positive for pedophilia.

Responses to female adults by the control subjects were noticeably higher as measured by phallometry and less so by finger pulse. As with the faking group, breathing rate was uninformative in this regard. In contrast
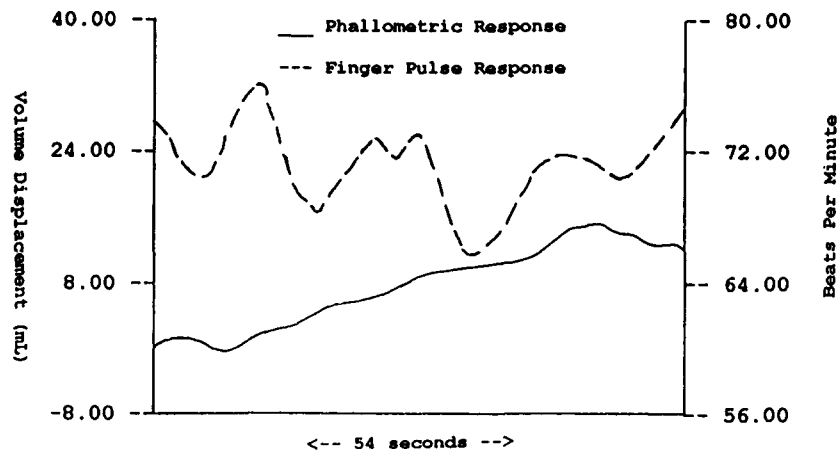

Fig. 2. Phallometric and finger pulse responses in a control subject.

to the fakers, the control subjects demonstrated no significant correlations between measures with the exception of a statistically significant negative correlation between the finger pulse and the breathing rate measures for responses to sexually neutral stimuli. Mean response and correlation data are summarized in Tables II and III. Figure 2 demonstrates, for an individual subject, the typical nonconcordance of phallometric and finger pulse responses to female adult stimuli.

DISCUSSION

The results of this study confirm earlier observations that the phallometric test of age preference is readily faked. This finding underscores the need to refine phallometric procedures in order to increase the validity of the test's stimuli, as well as to determine means to identify faked responses.

Nonetheless, phallometry may be less susceptible to distortion than are self-reports of sexual preferences. Subjects were instructed to fake in this study, while under conditions of assessment clients are not instructed to fake. However, the instructions given to fakers in this study were quite simple and the techniques suggested (and later reported by the fakers themselves) were the sort that most clients could discover on their own. We instructed the subjects in this study in order to *ensure* that they would attempt to fake a preference for female children. Telling them how to do this maximized the amount of usable data while providing a reasonable estimate of the worst-case scenario with respect to the test's susceptibility to faking.

Given our knowledge of the ease with which the test is faked, a critical task in the development of better test protocols is that of detection of faking. The results of this study confirm that psychophysiological measures of general arousal are not usable for individual diagnoses of erotic age preferences. However, there was some benefit in using finger pulse and, to a lesser degree, breathing rate as indicators of possible response manipulation. The results of this study were consistent with those reported elsewhere (Rosen *et al.*, 1975) that, under conditions of tension (such as a trial in which faking was occurring), heart rate is positively correlated with penile arousal and breathing becomes erratic. It could be argued that forensic and phallometric a sessments are intrinsically tension-producing. However, the client's experience of tension for these reasons would likely be constant throughout the procedure. Differential levels of tension would thus still provide useful information.

While a striking similarity was found in plotted test curves between penile volume and finger pulse in the responses of successful instructed fakers, only phallometric response was useful for individual diagnosis. This is consistent with an earlier observation that measures of general arousal are not usually specific enough to be used as measures of specific arousal and erotic arousal in particular (Zuckerman, 1971). Nonetheless, finger pulse appears to possess some utility as a method of detecting faking in phallometric procedures. Given the similarity of test plots, perhaps some measure of congruence could be used to generate a faking index.

Some difficulty was experienced in measuring breathing: the apparatus was susceptible to breakage because the lead wires were small and brittle. While the finger pulse device presented consistent and reliable measurement, the thermistor often deviated out of range and had to be manually recalibrated at points throughout a given test session. Most other studies reporting breathing rate data have used a strain gauge placed around the subject's chest. We chose to use another method because of the difficulties reported with regard to movement artifact with the strain gauge. Perhaps with further experimentation and practice, this method of measuring breathing will be more workable.

Results of most studies (e.g., Freund et al., 1988; Laws & Rubin, 1969) examining faking have consistently found that it is easier to suppress a response to an arousing stimulus than it is to produce arousal to one that is not. Those studies that questioned participants about their perceived ability to modify their responses have shown that subjects are good at estimating success in suppression, but they tend to overestimate the degree of success they had in producing arousal (Adams et al., 1992). Self-reports following testing in the faking group here indicate, however, that subjects were inclined to underestimate the success they had at faking a preference for female children. The levels of arousal to female children produced by the fakers in our study were almost-identical to those produced to female adults by the controls. Further, the fakers were able to suppress their responses to female adults to a level equivalent to that produced to female children by the controls. These results suggest that phallometric test subjects are able to control their penile responses under conditions of instructed faking.

It is often difficult to generalize results obtained with student controls to client populations. This is particularly so with sexual offenders, where the demand characteristics involved may be vastly different (Orne, 1962). Results obtained with students should thus be viewed cautiously. It is also important to recognize that faking does not necessarily indicate that the individual *is* paraphilic. Signs of faking should be used only to qualify or invalidate test results in which they are found.

One potential difficulty that might have influenced this study was the use of standardized scores rather than raw scores in the evaluation and comparison of penile responses. This has continued to be a topic of discussion in the literature (Barbaree & Mewhort, 1994; Earls, Quinsey, & Castonguay, 1987), as standardization of scores may increase or decrease the potential for a Type II error. In the present study, empirically determined cutting scores and evaluation procedures were used as means to control for error. Nonetheless, continual reevaluation and restandardization of procedures are crucial. Faking in phallometric testing will continue to be an important area of research in sexual offender assessment.

ACKNOWLEDGMENTS

REFERENCES

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.

Bancroft, J., Jones, H., & Pullan, G. (1966). A simple transducer for measuring penile erection, with comments on its use in the treatment of sexual offenders. *Behaviour Research & Therapy, 4,* 239-241.

Barbaree, H. E., & Mewhort, D. J. K. (1994). The effects of Z-score transformations on measures of relative erectile response strengths: A re-appraisal. *Behaviour Research & Therapy, 32,* 547-558.

Barlow, D., Becker, J., Leitenberg, H., & Agras, W. (1970). Technical note: A mechanical strain gauge for recording penile circumference change. *Journal of Applied Behavior Analysis, 3,* 73-76.

Earls, C. M., Quinsey, V. L., & Castonguay, L. G. (1987). A comparison of scoring methods in the measurement of penile circumference changes. *Archives of Sexual Behavior, 6,* 493-500.

Freund, K. (1957). Diagnostika homosexuality u muzu [*Diagnosing homosexuality in man*]. *Czechoslovakia Psychiatrica, 53,* 382-393.

Freund, K. (1963). A laboratory method for diagnosing predominance of homo- or hetero-erotic interest in the male. *Behaviour Research & Therapy, 1,* 85-93.

Freund, K., & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology, 57,* 1-6.

Freund, K., & Watson, R. (1990). Mapping the boundaries of courtship disorder. *Journal of Sex Research, 27,* 589-606.

Freund, K, & Watson, R. (1991). Assessment of the sensitivity and specificity of a phallometric test: An update of "Phallometric Diagnosis of Pedophilia." *Psychological Assessment, 3,* 254-260.

Freund, K., Sedlacek, F., & Knob, K. (1965). A simple transducer for mechanical measurement of the male genital. *Journal of Experimental Analysis of Behavior, 8,* 169-170.

Freund, K., Chan, S., & Coulthard, R. (1979). Phallometric diagnosis with 'non-admitters.' *Behaviour Research & Therapy, 17,* 451-457.

Freund, K., Watson, R., & Rienzo, D. (1988). Signs of feigning in the phallometric test. *Behaviour Research & Therapy, 26,* 105-112.

Geer, J. H. (1975 October). *Sexual functioning: Some data and speculations on psychophysiological assessment.* Paper presented at the Behavioral Assessment Conference, West Virginia University, Morgantown.

Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment, 4,* 502-511.

Henson, D., & Rubin, H. (1971). Voluntary control of eroticism. *Journal of Applied Behavioral Analysis, 4,* 37-44.

Kaine, A., Crim, M., & Mersereau, G. (1988). Faking sexual preference. *Canadian Journal of Psychiatry, 33,* 379-385.

Lalumiére, M. L., & Quinsey, V. L. (1993). The sensitivity of phallometric measures with rapists. *Annals of Sex Research, 6,* 123-138.

Lalumiére, M. L., & Quinsey, V. L. (1994). The discriminability of rapists from non-sex offenders using phallometric measures: A meta-analysis. *Criminal Justice and Behavior, 21,* 150-175.

Laws, D., & Holmen, M. (1978). Sexual response faking by pedophiles. *Criminal Justice & Behavior, 5,* 343-356.

Laws, D., & Rubin, H. (1969). Instructional control of an autonomic sexual response. *Journal of Applied Behavioral Analysis, 2,* 93-99.

Malcolm, B., Davidson, P., & Marshall, W. (1985). Control of penile tumescence: The effects of arousal level and stimulus content. *Behaviour Research & Therapy, 23,* 272-280.

Masters, W. H., Johnson, V. E., & Kolodny, R. C. (1994). *Heterosexuality.* New York: Harper Collins.

McAnulty, R. D., & Adams, H. E. (1991). Voluntary control of penile tumescence: Effects of an incentive and a signal detection task. *Journal of Sex Research, 28,* 557-577.

Orne, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776-783.

Oswald, P. J., & Cleary, P. J (1986). Effects of pelvic muscle tension and expectancy on general and specific indicators of sexual arousal. *Archives of Sexual Behavior, 15,* 247-260.

Quinsey, V., & Chaplin, T. (1988). Preventing faking in phallometric assessments of sexual preference. *Annals of the New York Academy of Sciences, 528,* 49-58.

Rosen, R. C., Shapiro, D., & Schwartz, G. E. (1975). Voluntary control of penile tumescence. *Psychosomatic Medicine, 37,* 479-483.

Rubin, H. B., & Henson, D. E. (1975). Voluntary enhancement of penile erection. *Bulletin of the Psychonomic Society, 6,* 158-160.

SPSS (1991). *Statistical package for the social sciences,* Version 4.0 for VAX/VMS.

Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin, 75,* 297-329.