

Phallometric testing with sexual offenders against female victims: Limits to its value

Doctoral Thesis, Queen's University¹

Yolanda Fernandez²

Advisor: William L. Marshall

Committee Members: Ronald Holden, Brian Butler, James Hillen, and Philip Firestone³

This article briefly summarizes a series of studies that explored issues related to the reliability (internal consistency and test-retest reliability) and criterion validity of phallometric testing with sexual offenders (incest offenders, extrafamilial child molesters, and rapists) against female victims. Three assessment sets were evaluated: (1) An Age-Gender set which presented slides of adults and children; (2) A Female Sexual Violence set which presented audio descriptions of consenting and forced sex between adults; and (3) A Child Sexual Violence set which presented audio descriptions of the sexual molestation of children by adults. A total of 280 incest offenders, 138 extrafamilial child molesters, and 139 rapists were included in the different analyses.

Introduction

In order to evaluate the adequacy of phallometry as a psychophysiological test, it is necessary to identify the standards against which the procedure is to be assessed. The value of a test is dependent on being standardized and can be shown to be both reliable and valid. At present, there is a paucity of research addressing the reliability of phallometric testing with sexual offenders and what little evidence is available is fraught with problems. Standards for acceptable levels of reliability vary according to a variety of factors, and there appears to be little agreement in the literature as to what exactly are satisfactory levels of reliability for phallometric testing.⁴

For criterion validity, phallometric results demonstrating differences between groups who are expected to differ need to be, within reason, consistently replicated. Extrafamilial child molesters have consistently been distinguished from non-offenders in terms of their responses to phallometric testing. However, the findings regarding rapists and incest offenders are less clear. While some studies have indicated that incest offenders respond more like non-offenders than do extrafamilial child molesters, other research has demonstrated no differences between extrafamilial child molesters and incest offenders.⁵ In addition, while child molesters have been distinguished from rapists using child stimuli, only one study⁶ has attempted to distinguish child molesters from rapists using adult rape stimuli, and

failed to find differences. Research addressing these inadequacies in the literature will have important implications for the interpretation and use of phallometric procedures in the assessment and treatment of sexual offenders.

Definition

Phallometric testing is the measurement of male erectile responses (Penile Plethysmography) during the presentation of various sexual and non-sexual stimuli. The stimuli are chosen to represent categories of sexual behaviour thought to be relevant to various offence patterns. Typically, these stimuli are categorized as either deviant (involving children, adolescents, or forced sex with adults) or appropriate (mutually consensual sex between adults) and the man's relative arousal to these categories of stimuli is calculated. In most cases, an individual must respond to an estimated 10% full erection or 3mm of change for the results to be considered clinically interpretable. Following conversion to standardized z-scores, deviance differentials are calculated which compare the various stimulus categories. It is these within subject comparisons that determine the presence of inappropriate sexual preference or interest.⁷

Method and procedure

Data were extracted from archival files on men currently serving federal sentences for offences against female victims (either adult or child). Based on their complete sexual offending history, offenders were classified according to their victim's age. Men with victims 18 years of age and older were classified as rapists, while men with victims younger than 14 years of age were classified as child molesters. The child molesters were further classified as either incest offenders or extrafamilial child molesters. Child molesters were assigned to the incest offender group if they had sexual contact with their daughter (biological, adopted or surrogate). In cases where the incest offender had multiple victims, all victims were required to be members of the nuclear family. Child molesters with both incestuous and extrafamilial victims were excluded, as were those offenders who had both child and adult victims. It has been recommended that a minimum response of 10% full

erection is required for a valid profile.⁷ Using this criterion, offenders with invalid response profiles were then removed from the analysis. Offenders were initially tested during their Intake Assessment at the Millhaven Assessment Unit and then re-assessed at a later time in their incarceration prior to treatment at either the Warkworth Sexual Behaviour Clinic in Warkworth Penitentiary or at the Regional Treatment Centre (Ontario) in Kingston Penitentiary. The mean test-retest interval was 6 months (range .5 to 25 months).

Phallometric results may be represented as raw scores in the form of either millimeter change in the circumference of the penis, voltage changes, or volume changes. Other scoring methods include transforming raw scores to percentage of full erection or standard scores (z-scores). In the present series of studies phallometric data were collected in raw form as mm-stretch and voltage changes but were converted to percentage of full erection for the internal consistency study and z-scores for the test-retest and criterion validity studies. Differential indices were calculated by subtracting the average z-score

response to an inappropriate category (for example, prepubescent women) from the average z-score response to an appropriate category (such as, the adult female). Differential indices greater than 0.0 were interpreted to indicate appropriate arousal, and values of 0.0 or less were interpreted to indicate inappropriate arousal.

Results

Internal consistency

The internal consistency results are summarized in Table 1. Internal consistency coefficients for all three assessment protocols (Age-Gender, Child Sexual Violence, and Female Sexual Violence), when assessed separately by offence type and stimulus category, were primarily at the moderate level with a few categories obtaining a high level of reliability. The results for the extrafamilial child molester group were somewhat less consistent with three categories of the Child Sexual Violence Assessment and one category of the Female Sexual Violence Assessment demonstrating unacceptably low internal consistency.

Table 1

Internal consistency reliability coefficients for the three assessments protocols by offence type								
Age-Gender Assessment								
Subject type		Women			Men			Neutral
		Adult	Prepubescent	Pubescent	Adult	Prepubescent	Pubescent	
Incest Offenders	(N = 143)	.90	.89	.95	.88	.90	.87	.92
Extrafamilial Child Molesters	(N = 67)	.87	.84	.84	.88	.85	.84	.87
Child Sexual Violence Assessment								
Subject type		Women		Child		Men		Neutral
		Adult	Passive	Coercive	Sexual Violence	Non-sexual Violence		
Incest Offenders	(N = 76)	.74	.78	.85	.85		.92	
Extrafamilial Child Molesters	(N = 31)	.83	.45	.65	.73		.63	
Subject type		Men		Child		Men		Neutral
		Adult	Passive	Coercive	Sexual Violence	Nonsexual Violence		
Incest Offenders	(N = 76)	.92	.94	.91	.75		.75	
Extrafamilial Child Molesters	(N = 31)	.98	.76	.43	.70		.40	
Female Sexual Violence Assessment								
Subject type		Consent		Rape		Violence		Neutral
		Partner	Narrator	Sexual	Anger	Robbery	Anger	
Incest Offenders	(N = 61)	.62	.86	.67	.68	.61	.84	.75
Extrafamilial Child Molesters	(N = 40)	.78	.79	.52	.85	.90	.80	.85
Rapists	(N = 139)	.80	.85	.71	.79	.71	.75	.69

Test-retest reliability

When the test-retest reliability of the Age-Gender and Female Sexual Violence Assessments was calculated separately for distinct offender groups, the correlation coefficients were generally less than acceptable (see Table 2). Only two of the stimulus categories from the Age-Gender Assessment (adult women and adult men) obtained acceptable (and then only moderate) levels of test-retest reliability. Among the differential deviance indices only three indices, from the gender preference analyses of the Age-Gender Assessment reached minimal acceptable levels of reliability. None of the stimulus categories and none of the differential deviance indices from the Female Sexual Violence Assessment reached acceptable levels of test-retest reliability. These data do not support the test-retest reliability of either assessment protocol for the two subjects types included in the present study.

Criterion validity based on contrasted groups

Incest offenders versus extrafamilial child molesters

The extrafamilial subjects were more deviant than the incest offenders in response to the Age-Gender Assessment (See Figure 1). In contrast, the incest offenders and the extrafamilial child molesters had similar response levels to the Child Sexual Violence Assessment (See Figure 2). Interestingly, however,

neither the responses to the Age-Gender Assessment nor the responses to the Child Sexual Violence Assessment, accurately identified group membership. While this is not surprising for the Child Sexual Violence Assessment, since both groups were equally deviant on this assessment set, it is somewhat puzzling that the responses to the Age-Gender Assessment did not predict group membership.

Within the incest offender group, subjects displayed more deviant arousal when assessed using the audiotaped presentation of stimuli (Child Sexual Violence) than when they were assessed using the slide presentation of stimuli (Age-Gender). In addition, more incest offenders were classified as deviant in response to the Child Sexual Violence Assessment than in response to the Age-Gender Assessment. In contrast, for the Extrafamilial child molesters there was no difference in terms of frequency of classification as deviant or non-deviant between the Child Sexual Violence and Age-Gender Assessments. Both assessment sets identified a substantial number of these subjects as deviant (See Table 3).

Rapists versus child molesters

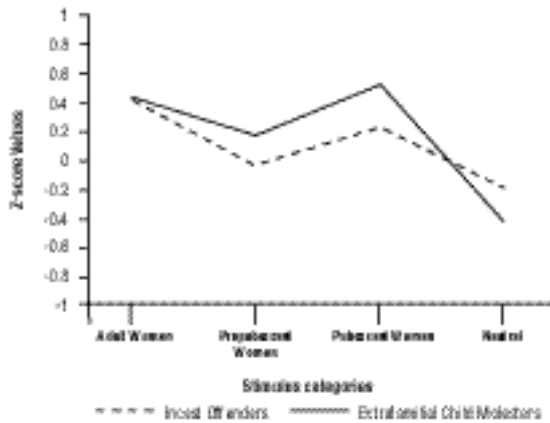
Surprisingly, the rapists did not respond more deviantly to the Female Sexual Violence Assessment than did either the incest offenders or the extrafamilial child molesters (See Figure 3). In addition,

Table 2

Test-retest reliability coefficients for the three assessment protocols by offence type							
		Age-Gender Assessment					
Subject type		Women			Men	Neutral	
		Adult	Prepubescent	Pubescent	Adult	Prepubescent	Pubescent
Extrafamilial							
Child Molesters	(N = 40)	.75	.18	.42	.74	.47	.49
		<i>Differential Deviance Indices</i>					
		Women–	Adult Women–	Child Female–			
		Male	Men	Male			
		.68	.79	.36			
		Adult Women–	Adult Women–	Adult Men–	Adult Men–		
		Prepub.	Pub.	Prepub.	Pub.		
		.56	.59	.74	.55		
		Female Sexual Violence Assessment					
		Consent	Consent	Rape	Rape	Violence	Violence
		Partner	Narrator	Sexual	Anger	Robbery	Anger
Rapists	(N = 51)	.48	.22	.32	.35	-.11	.11
		<i>Differential Deviance Indices</i>					
		Consent–	Consent–	Violence–			
		Rape	Violence	Rape			
		.56	.27	.16			

Figure 1

Mean responses of incest offenders and extrafamilial child molesters on the Age-Gender Assessment (women stimuli only)



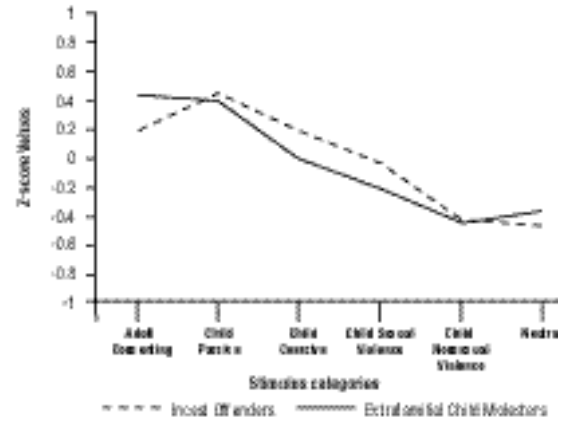
the Female Sexual Violence Assessment did not accurately predict group membership and did not classify as deviant more of the rapists than either the incest offenders or the extrafamilial child molesters.

Discussion

To date, the research on the psychometric properties of phallometric testing has been limited and inconsistent. Although previous internal consistency studies have often resulted in acceptable internal consistency coefficients, critics have argued that these studies have used unacceptably small samples sizes, and typically collapsed over stimulus categories and offender types, thereby possibly inflating the correlations.⁸ The present study, to

Figure 2

Mean responses of incest offenders and extrafamilial child molesters on the Child Sexual Violence Assessment (women stimuli only)



some degree, has put these criticisms to rest. Examining internal consistency levels for distinct stimulus categories and offender types resulted in satisfactory internal consistency to a limited degree. Acceptable internal consistency for each assessment set was limited to certain offender types. It should be noted, however, that the present study did not examine internal consistency for a group of "normals" or non-offender males and consequently, it is possible that internal consistency levels may be different for this group.

Previous studies of the test-retest reliability of phallometric testing have resulted in inconsistent results, although several studies reported acceptable levels of test-retest reliability. However, these studies

Table 3

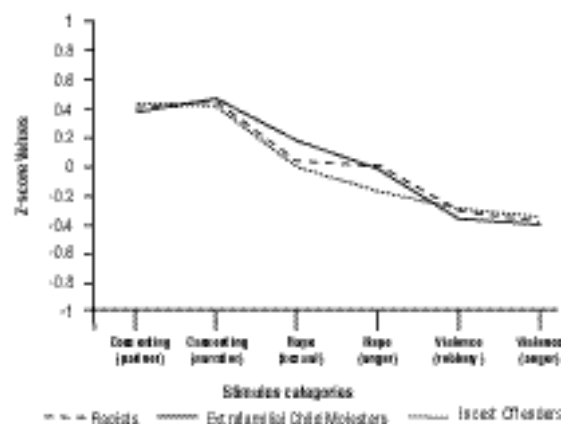
Comparison of deviance classification for Incest Offender and Extrafamilial Child Molester Groups

	Age-Gender Assessment n (%)	Child Sexual Violence Assessment n (%)
Incest Offenders		
Deviant	73(51)	57(75)
Non-deviant	70(49)	19(25)
Extrafamilial Child Molesters		
	Age-Gender Assessment n (%)	Child Sexual Violence Assessment n (%)
Incest Offenders		
Deviant	46(69)	20(67)
Non-deviant	21(31)	10(33)

Note: Values in parentheses are percent of total sample

Figure 3

Mean responses of incest offenders and extrafamilial child molesters on the Child Sexual Violence Assessment (women stimuli only)



typically examined the stability of phallometric testing over a relatively short period of time. This strategy is acceptable when the research to be conducted may not involve repeated testing or when repeated testing is to occur over a brief time span. However, the present results indicate that the test-retest reliability of at least two stimulus sets (Age-Gender and Female Sexual Violence) is clearly unacceptable when the testing period is extended to several months. Thus when long-term stability is needed, phallometric testing does not meet acceptable criteria. Obviously, however, the present finding needs to be replicated and future research studies will need to include the Child Sexual Violence Assessment, as this was not examined in the present series of studies.

The ability of phallometric test results to distinguish between groups of sexual offenders is directly relevant to the presumed role of sexual motivation in theories of sexual offending. Early theoretical perspectives suggested that sexual offenders enacted their deviant behaviours because they had developed conditioned arousal to the persons or actions involved in their deviant acts prior to enacting deviant behaviour.⁹ These theoreticians suggested various ways by which these deviant sexual preferences were acquired (for example, unexpectedly seeing a child while sexually aroused or while masturbating) but they were all said to result from conditioning processes that associated the stimuli with sexual arousal. The early theorists appeared to believe that *all* sexual offenders were driven by these acquired sexual preferences. Later conditioning theorists however, have suggested that only some sexual offenders have acquired deviant preferences.¹⁰ In any event, conditioning theories, and indeed most of the more comprehensive etiological accounts of sexual offending suggest that a substantial number of these men will display deviant sexual arousal at phallometric testing. In addition, the current popular use of phallometry certainly suggests that an underlying assumption operating in clinical settings is that sexual preferences function as a trait and is not remarkably influenced by circumstance. The exception to this rule, however, appears to be incest offenders, who are generally accepted as being more opportunistic (or as regressing to more juvenile modes of responding) than motivated by sexual preferences for children.

In the present study, the responses of extrafamilial child molesters to the slides of unfamiliar children suggest this group may have a more generalized interest in children than do the incest offenders. However, the incest offenders responded more deviantly to the audio presentation of stimuli. Perhaps the generally accepted notion that incest

offenders are opportunistic offenders rather than being motivated by a sexual attraction to children may be in error. It is possible that incest offenders do not have a broadly generalized sexual attraction to children, but instead are specifically sexually attracted to their own victims. In that case the results suggest that incest offenders are more likely to respond to children in a particular set of circumstances (within the home and under their authority) but are not likely to feel sexually attracted to children in general (for example, children playing on a playground). Extrafamilial offenders, who typically molest several different children, would by this same reasoning, be expected to have expanded or generalized their responses to the general class of children and might be expected to display sexual arousal toward children independent of circumstances. The results of the present series of studies then suggests that the distinction between extrafamilial child molesters and incest offenders, in terms of whether or not their offending is sexually motivated may have been misdirected. Instead perhaps theories of sexual offending against children should focus on different offender characteristics and perhaps situational features that lead to the offending behaviour.

Unfortunately, the data on the importance of sexual motivation for rapists, as assessed using the Female Sexual Violence Assessment, is considerably less convincing. The results of this study demonstrated that the responses to the Female Sexual Violence Assessment did not provide any information unique to rapists. Rapists and child molesters did not respond differently to the assessment and neither group demonstrated a preference for rape stimuli. These data do not support the sexual preference hypothesis, as it has been applied in theories about rapists. Consequently, the issue of deviant arousal in rapists and the processes by which rapists acquire deviant arousal may not be relevant to understanding the motivation of rapists to commit their offences (in so far as phallometry is considered an index of deviant arousal). Instead theorists may need to consider explanations of rape that take into account factors apart from the sexual motivation and, in particular, sexual preferences. In fact, many theories of rape propose that a need to exert power over and to humiliate women is the prime motivation for rape.¹¹ Another approach may be for theorists to consider that sexual preferences may not function as a trait but rather may be dependant upon current internal state (high arousal, negative mood states, intoxication) or novelty of the stimuli (never seen before). Some researchers have already focussed their efforts on the effects of such factors on deviant sexual arousal as assessed by phallometric testing.

Finally, the failure to detect deviant arousal in rapists, and the findings concerning arousal to children among incest offenders and non-familial child molesters (approximately 50% of the incest offenders and 70% of the extrafamilial child molesters displayed deviant arousal to the Age-Gender Assessment), casts doubt on the external validity of phallometric testing. These results

suggest a poor match between deviant sexual preferences revealed at phallometric evaluations and a history of actual deviant behaviour. Although the present results suggest this as a possibility, a more rigorous study is needed that includes an examination of responses and sexual history of both identified offenders and non-offending men, before this issues can be said to be addressed satisfactorily. ■

¹ Abstract from Fernandez, Y. M. (2000) *Phallometric testing with sexual offenders: Limits to its value*. Doctoral Thesis, Kingston, ON: Queen's University.

² 340 Laurier Avenue West, Ottawa, Ontario, K1A 0P9.

³ Philip Firestone, Ottawa University, Ottawa, Ontario.

⁴ Barbaree, H. E. (1990). Stimulus control of sexual arousal. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 115 - 142). New York, NY: Plenum Press.

⁵ Looman, J. (2000). *Sexual arousal in rapists and child molesters*. Unpublished Doctoral thesis. Kingston, ON: Queen's University.

⁶ Marshall, W. L., and Eccles, A. (1991). Issues in clinical practice with sex offenders. *Journal of Interpersonal Violence*, 6, 68-93.

⁷ 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, 123-174.

⁸ Abel, G. G., and Blanchard, E. G. (1974). The role of fantasy in the treatment of sexual deviation. *Archives of General Psychiatry*, 30, 467-475.

⁹ Laws, D. R., and Marshall, W. L. (1990). Aconditioning theory of the etiology and maintenance of deviant sexual preferences and behavior. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 209-229). New York, NY: Plenum Press.

¹⁰ Darke, J. L. (1990). Sexual aggression: Achieving power through humiliation. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds.), *Handbook of sexual assault: Issues, theories and treatment of the offender* (pp. 55-72). New York, NY: Plenum Press.

¹¹ Russell, D. E. H. (1984). *Sexual Exploitation: Rape, child sexual abuse and workplace harassment*. Thousand Oaks, CA: Sage Publications.

Research in Brief

The Research Branch of Correctional Service of Canada (CSC) has developed a publication entitled Research in Brief. This is a series of one-page summaries of research reports specifically designed to more broadly disseminate correctional knowledge to others.

Volume 1, R-01 to R-100, provides summaries covering the first one hundred research reports published to date. Volume 2, R-101+ is being prepared and will be made available shortly.

Copies of Research in Brief and the full report are available via the CSC web site.