

## Assessment of the Sensitivity and Specificity of a Phallometric Test: An Update of Phallometric Diagnosis of Pedophilia

Kurt Freund and Robin J. Watson  
Department of Behavioural Sexology  
Clarke Institute of Psychiatry  
Toronto, Ontario, Canada

The specificity and sensitivity of the phallometric test of an erotic preference for minors was assessed. The specificity was determined to be 96.9% if using a group of sex offenders against female adults and 80.6% if using a group of paid volunteers. Test results of 27 sex offenders against at least 2 female children each and of 22 offenders against at least 2 male minors each (either against children or against pubescents, but not against both), demonstrated sensitivities of 78.2% for heterosexual pedophiles and 88.6% for homosexual pedophiles or hebephiles. From these test sensitivities, the percentage of subjects preferring minors in a group of offenders against only 1 female child each, a group of offenders against at least 1 female child *and* at least 1 female pubescent each, and a group of offenders against only 1 male minor each (child or pubescent) were determined as being 44.5%, 74.6%, and 86.7%, respectively.

The differential diagnosis between the various erotic preferences can be established by means of the phallometric test, which uses continuous recording of penile volume changes while a subject views potentially arousing pictures on screens or listens to such taped narratives (Abel, Becker, Murphy, & Flanagan, 1981). The current study was limited to the phallometric test of pedophilia and homosexual hebephilia as it is used in the differential diagnosis between pedophilia or homosexual hebephilia and gynephilia (Freund, Diamant, & Pinkava, 1958; Quinsey, Steinman, Bergersen, & Holmes, 1975).

The revised third edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*; American Psychiatric Association, 1987) lists the essential feature of pedophilia as "recurrent, intense, sexual urges and sexually arousing fantasies, of at least six months' duration, involving sexual activity with a prepubescent child" (p. 284).

The current study defines *pedophilia* as an erotic preference for children (Freund, McKnight, Cibiri, & Langevin, 1972). Analogously, *hebephilia* is defined as an erotic preference for pubescents; *gynephilia*, as a preference for female adults; and *androphilia*, as a preference for male adults. In the following, an erotic preference for children or pubescents will be referred to as a preference for *minors*, which is not identical with the legal meaning of the term. Our definitions of pedophilia and hebephilia also deviate slightly from the supposed age limit of childhood indicated in *DSM-III-R* as 13 years. We have been using a

more conservative estimate of 11 years. During the past 2 decades, variants of this test have been administered in psychological labs at various locations for the purpose of individual diagnosis.

Wherever the phallometric test of pedophilia and homosexual hebephilia is used, it uses stimuli that represent children and adults by means of either or both slides or filmstrips shown on a screen or verbal narratives. There are, however, great differences among the various labs in stimulus presentation, type of penile sensor, and evaluation of test results. A number of studies have endeavored to assess the diagnostic potential of the test. Examples of such studies are those by Frenzel and Lang (1989), Marshall, Barbaree, and Butt (1988), and Marshall, Barbaree, and Christophe (1986).

Every medical diagnostic lab has the obligation to measure the specificity and sensitivity of the test procedures it is using as rigorously as is possible at the time. This is also a necessity for psychophysiological testing labs. However, according to our knowledge, no such rigorous assessment has been carried out for the phallometric test of pedophilia and homosexual hebephilia. The current study was intended to fill this gap.

### Subject Selection for Assessment of Specificity

The specificity of a biological or psychological diagnostic procedure is assessed by determining the proportion of positive test outcomes in a group of subjects supposed to not possess the attribute the test is meant to diagnose; this represents the control group. However, the selection of control subjects for a psychological test must also ensure that the demand characteristics (Orne, 1962) of their test situation are as similar as possible to those of the experimental group. In the present case, the experimental group consisted of sex offenders whose victims were minors, and who claimed to be gynephilic, as almost all of them do.

---

This research was supported in part by Grant MA-9134 from the Medical Research Council of Canada.

We thank Michael Kuban for editorial assistance, Cathy Spegg for statistical advice, and George Whiting for programming and design of the phallometric software.

Correspondence concerning this article should be addressed to Kurt Freund, Department of Behavioural Sexology, Clarke Institute of Psychiatry, 250 College Street, Toronto, Ontario, Canada, M5T 1R8.

The diagnostic potential of the test is much less satisfactory with sex offenders against minors who do not admit to an erotic preference for this age bracket than with offenders who admit to such a preference (Freund, 1967). The reason for this is that a substantial proportion of subjects are able to influence the test outcome according to the demand situation, but this is not true for the mentally handicapped (Freund & Blanchard, 1989). A number of objective signs of manipulation of test responses have been detected and were included in the standard evaluation of the test result (see below). However, a substantial amount of undetected faking remains (Freund, Watson, & Rienzo, 1988).

In a recent validation study of the phallometric method (Freund & Blanchard, 1989), of which the present study is a follow up, sex offenders against female adults were used as control subjects because it was supposed that, although never charged for involving a child in their sexual activities, this type of offender has a particularly strong interest not to appear pedophilic or hebephilic. Before the present investigation, the validity of the supposition that sex offenders against female adults are a more suitable control group than nonoffender volunteers had not been tested. This issue was addressed by Part 1 of the current study.

### Selection of Index Cases

The sensitivity of a biological lab test is measured by assessing the proportion of positive test results in a group of individuals supposed to possess the attribute that the test is meant to diagnose; this group represents the *index cases*. For the assessment of the diagnostic potential of the phallometric test of an erotic preference for minors, index cases would be subjects supposed to have such a preference.

It is most likely that such individuals will be found among sex offenders against minors and will be particularly frequent among offenders against more than one minor. Incest offenders and exhibitionists are exceptions (Freund & Watson, 1990; Quinsey, Chaplin, & Carrigan, 1979), as are those individuals who erotically prefer pubescent girls but are not pedophilic; the phallometric test does not appear to be capable of diagnosing heterosexual hebephilia proper. Thus, multivictim offenders may be the best suited index cases. On the other hand, the supposition that there are sex offenders against minors who are gynephilic may hold true not only for such special groups as incestuous offenders, exhibitionists, and possibly sadists, but for all sex offenders against female children, including the multivictim offenders (although to a lesser extent).

Clinical experience has led to the preliminary estimate that approximately 90% of multivictim nonincest sex offenders against female children are true pedophiles, whereas the corresponding percentage in multivictim sex offenders against at least two male minors may be 95%. The choice of a somewhat larger proportion for offenders whose victims were male minors was supported by phallometric studies, which showed that on the gynephile's scale of erotic preferences the male minor occupies a position even lower than that of the female child (Freund, McKnight, Cibiri, & Langevin, 1972; Freund, Watson, & Rienzo, 1989).

The selection of index cases, however, has to be further re-

stricted; it is seldom necessary to administer the test for the reason of diagnosis in the rare cases of men who readily admit strong erotic attraction to minors, therefore only "nonadmitters" should be selected. Nonadmitters are those who deny an erotic preference for minors and claim to be gynephilic (the test has not yet been standardized for a differential diagnosis between a preference for minors and androphilia).

## Part 1

### Method

#### Subjects

There were two groups of nonpsychotic men aged 18 to 55 years (inclusive) who claimed to be gynephilic. The first group consisted of 50 paid volunteers. These men were recruited from a government placement office for the unemployed and from community colleges (mean age = 26.1 years,  $SD = 6.8$  years; median educational level = 12 grades completed, but no university). The majority of the unemployed and approximately one third of the students of the community colleges were unskilled laborers (the Hollingshead-Redlich, 1958, index for occupational status was not available). Approximately one fifth of the unemployed and one quarter of the community college students were of Asian or African ethnicity. The second group consisted of 41 sex offenders who had been charged with assaulting female adults and who did not have any known sex offense against a minor. The known paraphilic behaviors displayed by this offender group were as follows: 8 individuals demonstrated sadism, 23 individuals had committed rape, 17 individuals had exposed; 2 individuals demonstrated each of transvestism, fetishism, and toucheurism; and 1 individual displayed voyeurism and made obscene telephone calls. Because the majority of individuals in this group displayed more than one such paraphilic behavior, there is overlap in this listing. The group's mean age was 25.8 years ( $SD = 5.8$  years), and its median education level was more than 8 grades completed but less than 12 (Hollingshead-Redlich index: 7—unskilled labor). Almost all of the individuals were White.

All offenders, those against women as well as those against minors, were referred to us by psychiatrists who were assessing them either for lawyers (as was the case for the majority of these subjects) or within the prison system. Exhibitionists who did not express any other paraphilic activity according to which they could be identified as gynephiles were not included; there is the possibility that such offenders may actually have preferred minors (Freund & Watson, 1990).

#### Verbal Procedure

Assessment of gynephilia in the subjects included in this part of the study was by means of the Erotic Preferences Examination Scheme (EPES; Freund, 1965), a multiple-choice questionnaire used routinely in our department.

#### Phallometric Procedure

Subjects in both groups were told that the test would establish whether they were attracted to children in addition to adults. The phallometric test of pedophilia and homosexual hebephilia was administered in two sessions. However, a subject diagnosed in Session 1 as erotically preferring minors was not required to undergo Session 2. As in our earlier studies, the original volumetric penile sensor was used (Freund, Sedlacek, & Knob, 1965).

In Session 1, color filmclips of nude individuals, including both genders, of four age categories were presented: 5- to 8-year-old children, 8- to 11-year-old children, pubescents, and physically mature

persons. Additionally, sexually neutral filmclips (i.e., landscape scenes) were shown. This test session consisted of 3 blocks of 9 trials (27 in total), each of 28 s duration. In the 1989 study, every trial consisted of two consecutive filmstrips of 14 s duration that showed either different individuals of the same gender–age category or landscapes. These stimuli were presented on one screen. In the current study, three screens simultaneously presented different persons of the same gender–age category. In this way, altogether six persons of the same gender–age category were shown in each trial, during which a taped narrative describing involvement of the depicted persons in nonsexual activities (such as swimming) accompanied the filmstrips. The narratives stressed characteristic features of the depicted persons' body shape, and the landscape scenes had associated sexually neutral descriptions. The depictions of the eight gender–age categories and the landscapes were presented in fixed random succession with the exception that direct succession of trials of the same gender–age category was not allowed.

In Session 2, slides were presented showing nude 8- to 11-year-old girls and boys, physically mature persons of both genders, and landscapes. In each trial, nine slides depicting three different views of three different persons of the same gender–age category were presented such that each of three screens showed three slides—front view, rear view, and genital region—in fixed random succession. No two slides with the same type of view were shown at the same time. Each gender–age category was presented once in each of five blocks arranged in fixed random order with the exception that direct succession of two trials of the same gender–age category was not allowed. Each trial was accompanied by a narrative depicting a person of the presented gender–age category in sexual interaction with the subject.

### Measurement

For each trial, penile volume changes (in mL) were measured in two ways: (a) a *D*-type score, the largest deviation from initial value, and (b) an *A*-type score, the area under the plotted response curve. The raw *D*-type scores for each separate session were converted to standard scores derived from each subject's own *D*-type data only (Freund et al., 1972), and the same operation was carried out with his *A*-type scores. These *D*-type and *A*-type standard scores were subsequently combined according to the formula  $(z_D + z_A)/2$ . The area under the curve *A* and maximal deviation from baseline *D* are highly correlated and therefore any response with an uncharacteristic relationship between these two measures is likely to reflect an irregularity like, for instance, body movement. The combined score lowers the impact of such irregularities in the evaluation of test results.

### Evaluation of Results

To avoid inclusion of test sessions in which there was virtually no response to the stimuli, we determined a minimal value of response output necessary to consider a test session valid. This *output index* (hereinafter O.I.) was the mean of the three highest responses in raw *D* scores, excluding responses to neutrals. In the earlier study (Freund & Blanchard, 1989), the minimum acceptable value of this index was 0.5 mL for Session 1 and 1 mL for Session 2. In the present study, a value of 1.0 mL was necessary for each session. In a test session in which the subject had responded most to sexually neutral stimuli, a minimal O.I. of 1.5 mL was required.<sup>1</sup>

### Statistical Analysis

The comparison of proportions of (supposedly erroneous) diagnoses of an erotic preference for minors was by two-tailed *t* test of proportions with correction for small samples. The comparison of the two

groups in regard to education was by a nonparametric test (Mann-Whitney *U* test). The difference between the groups in regard to mean age was too small to warrant statistical comparison.

## Results

The results pertaining to the specificity of the test were as follows: A preference for minors was diagnosed in 19.4% of the volunteers; however, the offenders against female adults (for whom the demand characteristics of the test situation were similar to those for the offenders against minors) received such a diagnosis in only 3.1% of the cases. The *t* test of proportions showed this difference to be significant:  $t(66) = 2.085$ ,  $p < .05$ . There was no significant difference in regard to education.

### Part 2

Part 2 of the study attempted to assess, in more detail than the earlier similar study (Freund & Blanchard, 1989), the degree of sensitivity of the phallometric procedure with groups of multivictim offenders against minors. These measures of sensitivity were then used to estimate the proportions of individuals with an erotic preference for minors in other groups in which the presence of such individuals may be supposed to be smaller. This operation was based on the supposition that the sensitivity of the test remains constant regardless of whether the test is administered to single-victim offenders against minors or to multivictim offenders against minors. However, only two of the groups of multivictim offenders were large enough to be used in assessing the test's sensitivity: the offenders against female children and the offenders against male minors (i.e., the groups of offenders against male children and that of offenders against male pubescents combined).

### Method

#### Subjects

A total of 147 nonpsychotic sex offenders against minors, aged 18 to 55 years inclusive, were selected for the study. All were nonadmitters and were of at least dull average intelligence. Table 1 shows group size, mean, and standard deviation of age, median level of education, median occupational status, lowest age bracket of victims, the numbers of individuals who also engaged in incestuous activities, and the number of offenders who had been included in the above-mentioned earlier

<sup>1</sup> Further evaluation was identical with that described in the earlier article (Freund & Blanchard, 1989), except for a few later refinements in the evaluation of results: When the highest response was to pubescent girls or if in both test sessions responses to adults and children were equal, the present study diagnosed a preference for minors and response suppression. This had not been so in the earlier study. A further addition was in regard to internal validity indicators (detection of "faking"): If in Session 1 a subject responded at least 0.25 z less to the stimuli depicting girls 8 to 10 years old than he did to girls 5 to 8 years old, or if he showed an analogous pattern in regard to male individuals, this was also evaluated in the present study as response suppression. More detailed information, in particular about gender and age preference indices, is available on request (since Freund & Blanchard, 1989, minor changes have been instituted, mainly in regard to cutoff points).

investigation of the diagnostic power of the test (Freund & Blanchard, 1989). Almost all were White.

Not included in the study proper were (a) offenders who exclusively victimized female children for whom they were in a father role; (b) individuals of maximally borderline intelligence; (c) full or partial admitters (see below); and (d) individuals who claimed to be androphilic.

**Procedures**

Nonadmitters were selected by the following two questions embedded in the EPES:

1. *When lying in bed and sexually aroused, do you imagine or fantasize about intimately touching (a) boys or girls up to 10 years old, (b) boys or girls 12 to 15 years old, (c) boys or girls in both age ranges, or (d) none of the above.*
2. *When lying in bed and sexually aroused, do you imagine or fantasize about intimately touching women 17 years old and older more often than touching children or teenagers 15 years old or younger? (a) yes, (b) no.*

Only those offenders against minors who endorsed *d* for Question 1 and *a* for Question 2 were selected. *Partial admitters* were those who chose *a*, *b*, or *c* for Question 1 and *a* for Question 2; *full admitters* were those who chose *a*, *b*, or *c*, for Question 1 and *b* for Question 2. Those few individuals who were not able to read and write were administered these EPES questions orally.

The phallometric test used for diagnosing an erotic preference for minors in Part 2 of the study was identical with that described in Part 1.

**Statistical Comparisons**

The comparisons in regard to diagnostic sensitivity were by two-tailed *t* test of proportion with adjustment for small samples. The comparison of the groups in regard to mean age was by analysis of variance and Tukey's honestly significant difference tests; the comparisons in regard to education and occupational status were by corresponding nonparametric tests (i.e., Kruskal-Wallis and Mann-Whitney *U* test).

**Pertinent Formulas**

The measure of sensitivity of the test was calculated separately for multivictim offenders against female children and for such offenders against male minors, according to the formula

$$sens = a/(b \times c),$$

in which *a* is the number of individuals diagnosed as having an erotic preference for minors, *b* is the total number of individuals who received a valid diagnosis (undetermined age preference included), and *c* is the clinical estimate of the number of individuals in the group (given earlier as 90% for multivictim offenders against female individuals and 95% for multivictim offenders against male individuals). Subjects whose test results indicated successful faking or whose O.I. was too low in both test sessions were excluded from this assessment.

Calculation of what the expected proportion (*c'*) of pedophiles and homosexual hebephiles might be in the other groups requires a reworking of the formula as follows:

$$c' = a/(b \times sens).$$

In this expression, *sens* is the sensitivity calculated for the corresponding group of multivictim offenders. This formula was applied to the following groups: (a) single-victim offenders against female children, (b) offenders against at least one female child and at least one female pubescent, and (c) single-victim offenders against male minors.

**Table 1**  
**Group Characteristics**

Characteristic	Offenders against female minors			Offenders against male minors			Offenders against minors of both genders			Gynephilic offenders (n = 41)
	Single (n = 29)	Multiple (n = 33)	Child + pubescent (n = 14)	Single (n = 18)	Multiple (n = 25)	Child + pubescent (n = 8)	Multiple (n = 12)	Child + pubescent (n = 8)	Multiple (n = 12)	
Age (years)*										
M	32.4 <sub>a,b,c</sub>	36.6 <sub>a,b</sub>	40.3 <sub>a,c</sub>	29.3 <sub>a,c</sub>	34.0 <sub>a,b,c</sub>	36.6 <sub>a,b,c</sub>	37.6 <sub>a,b</sub>	34.0 <sub>a,b,c</sub>	37.6 <sub>a,b</sub>	25.8 <sub>b,c</sub>
SD	10.0	9.6	9.9	10.4	10.0	9.3	7.5	10.6	7.5	5.8
Median education	8 < Ed < 12	8 < Ed < 12	8 < Ed < 12	Ed = 12	8 < Ed < 12	8 < Ed < 12	8 < Ed < 12	8 < Ed < 12	8 < Ed < 12	8 < Ed < 12
Median occupation**	USL <sub>a,b,c</sub>	USL <sub>a,c</sub>	SSL <sub>a,b</sub>	USL <sub>a,b,c</sub>	SL <sub>b</sub>	USL <sub>a,b,c</sub>	USL <sub>a,b,c</sub>	USL <sub>a,b,c</sub>	USL <sub>a,b,c</sub>	USL <sub>c</sub>
Youngest victim										
<6 years old	10	10	5	4	3	1	4	4	4	—
6-11 years old	19	23	9	6	11	7	8	4	8	—
12-15 years old	—	—	—	8	11	—	—	—	—	—
Deny event(s)	9	9	2	6	3	2	4	1	4	—
Number of incest (also deny)	—	13 (3)	9 (1)	—	2	2	5 (2)	5 (1)	5 (2)	—
Number in F & B, 1989	20	10	5	12	12	5	5	2	5	32

*Note.* Pairs of figures sharing like subscripts are not significantly different at *p* < .05. Single = victimized only 1 minor; Multiple = victimized at least 2 children (or at least 2 pubescent boys); Child + pubescent = victimized at least 1 child and at least 1 pubescent; gynephilic offenders = offenders against female adults with no offenses against minors; USL = unskilled laborer; SSL = semiskilled laborer; SL = skilled laborer; Deny event(s) = deny even that offense(s) occurred; Number of incest = number of offenders who also engaged in incestuous relationships (also deny = deny that offenses) occurred; Number in F & B, 1989 = number of subjects included previously in Freund & Blanchard (1989).  
\*F(6, 65) = 7.66, *p* < .01.  
\*\* Hollingshead & Redlich (1958),  $\chi^2 = 14.64$ , *N* = 170, *p* < .03.

## Results

Table 2 lists the diagnostic results, which include, only for clinical illustration, the results of the full and partial admitters and of the individuals with borderline or lesser intelligence (so diagnosed by the institutions or agencies that supervised them). The results of the group of 41 sex offenders against female adults that was described in Part 1 of this study are also included in Table 2.

The phallometric test diagnosed an erotic preference for children in 78.2% of the clinically estimated number of pedophiles among the multivictim offenders against female children. The test diagnosed an erotic preference for minors in 88.6% of the clinically estimated number of individuals preferring minors, among the multivictim offenders against male minors (i.e., the groups of multivictim offenders against male children combined with that of such offenders against male pubescents). The test results were not valid for 20.7% of the single-victim offenders against female children, 18.2% of the multivictim offenders against female children, 14.3% of the offenders against female children and pubescents, 27.7% of the single-victim offenders, and 24% of the multivictim offenders against male minors.

The estimates for the proportions of individuals with an erotic preference for minors among other groups of offenders, arrived at by means of the formula given here earlier, were as follows: approximately 45% were pedophiles among the single-victim offenders against female children, 75% were pedophiles or hebephiles among offenders against at least one female child

and at least one female pubescent; and 87% were pedophiles or hebephiles among single-victim offenders against either male children or pubescents.

The proportion of individuals among the single-victim offenders against female children phallometrically diagnosed as preferring minors was significantly smaller than that in the group of single-victim offenders against male children,  $t(34) = 2.43, p < .02$ . It was also significantly smaller than the proportion of individuals erotically preferring minors among the multivictim offenders against female children,  $t(48) = 2.52, p < .02$ . The results of the comparisons of groups according to age, education, and occupational status are shown in Table 2.

## Discussion

The main task of this study was assessment of the diagnostic power of the phallometric test for pedophilia and homosexual hebephilia. In contrast to *DSM-III-R*, this study defines these paraphilias as erotic preferences. This has the advantage, among others, that erotic preferences can be assessed by methods that, compared with self-reports, are less dependent on patients' willingness to disclose their erotic likes and dislikes.

Part 1 of the study supported the notion that when validating this test, offenders against female adults (who have a genuine interest not to appear as also substantially attracted to minors) are more suitable as control subjects than unselected paid volunteers. The volunteers were, in 19.4% of the cases, diagnosed as erotically preferring minors, whereas there were only 3.1% of such diagnoses in the offenders against female adults.

Table 2  
*Phallometric Diagnosis*

Type of offender	<i>n</i>	Low output index	Erotic preference for minors	Gynephilia	Undetermined age preference	Documented faking
Against female minors						
Single	29	1	8	13	2	5
Multiple	33	3	19	6	2	3
Child and pubescent	14	0	7	4	1	2
Mentally handicapped	8	0	7	0	1	0
Full admitters	10	0	7	2	1	0
Partial admitters	17	3	12	1	0	1
Against male minors						
Single	18	2	10	2	1	3
Multiple	25	3	16	2	1	3
Child and pubescent	8	0	5	1	0	2
Mentally handicapped	17	1	12	1	1	2
Full admitters	18	1	16	0	1	0
Partial admitters	11	0	9	1	0	1
Offenders against minors of both genders						
Multiple	12	2	5	1	1	3
Child and pubescent	8	1	3	1	3	0
Mentally handicapped	7	1	4	0	0	2
Full admitters	5	0	4	1	0	0
Partial admitters	5	1	2	1	0	1
Gynephilic offenders	41	3	1	29	2	6

*Note.* Single = victimized only 1 minor, multiple = victimized at least 2 children (or at least 2 pubescent boys); child and pubescent = victimized at least 1 child under age 12 and at least 1 pubescent; mentally handicapped = individuals of maximally borderline intelligence, full/partial admitters—see Method; gynephilic offenders = offenders against female adults with no offense(s) against minors.

The most plausible explanation for the disparity between the two gynephilic groups appears to be that they differ in regard to the demand characteristics of the test situation. Volunteering nonoffenders are most likely to be not very concerned about their test outcome and may not pay enough attention to the presented stimuli. In contrast to these volunteers, the sex offenders against female adults may apply considerably more effort in suppressing responses to minors—gynephilic men have been shown to also respond to some extent to female children (Freund et al., 1972).

If the difference between the two groups in regard to the diagnosis of an erotic preference for minors was due to the unequal demand situation, it may not be possible to find a fully adequate gynephilic control group because its demand situation will always differ from that of pedophilic or hebephilic nonadmitters, who may exert a particular effort to hide their pedophilia. A rigorous test of this conjecture, however, has not yet been attempted.

Another possible explanation for the differing results in regard to the test's specificity is that sexually normal male individuals are erotically more interested in children than are offenders against female adults. However, this is unlikely because, according to Abel, Becker, Cunningham-Rathner, Mittelman, and Rouleau (1988), there is a high proportion of offenders against female children (44%) among rapists.

Part 2 of the study assessed the sensitivity of the phallometric test of pedophilia and homosexual hebephilia, using offenders against at least two minors as index cases. The assessed sensitivity for heterosexual pedophilia was 78.2%; for an erotic preference for male minors, it was 88.6%. These results were then used in estimating the proportions of pedophiles or homosexual hebephiles in groups of offenders against only one female child (44.5%), against at least one female child *and* at least one female pubescent (74.6%), or against only one male minor (86.7%). These estimates are preliminary because of the small number of subjects.

The differentiation, however, between offenders against only one minor and offenders against two or more minors, used in this and earlier studies, may impose a dichotomy that in reality does not exist, and can be only approached to some (unknown) degree because most sex offenders against minors commit many more offenses than those for which they are apprehended, as demonstrated by Abel et al. (1988).

Although there were differences in age, education, and social status between the groups, the distribution of these differences was such that there was little reason to suppose they could have substantially distorted the described assessments. The findings, however, that sex offenders against female adults tend to be younger (and have a lower occupational status) than the offenders against minors, may reflect a possible connection of some forms of rape with particularly high sex drive. The tendency of the offenders against male minors to have a higher occupational status may reflect the larger proportion of true pedophiles or hebephiles among them, whereas nonpedophilic sexual offenses against female children may be more frequent in a lower social stratum. These conjectures, however, will have to be empirically validated.

The results of the current study have to be evaluated in the context of relevant, more general epidemiological aspects. Bio-

logical testing of a population for a disease (e.g., tuberculosis) consists essentially of two steps. First, a *screening* test is performed, which narrows the sample to a group of suspect cases—those with a positive result. In these cases, the screening test is followed up by a more reliable but more time-consuming or invasive diagnostic procedure that will confirm or disconfirm the diagnostic hypothesis arrived at from the screening test.

In the process of diagnosing an erotic preference for minors, a sex offense (by an adult) against a minor could be regarded as a positive outcome of a screening test for a corresponding erotic preference. It appears also warranted to perceive the number of times the offender was apprehended for involving a minor in his sexual activities as the number of screening tests he has had—all with a positive outcome. If this supposition is accepted, the phallometric procedure can be conceived of as a *confirmatory* test of those individuals whose screening tests had a positive result.

Griner, Mayewski, Mushlin, and Greenland (1981) and Griner and Glaser (1982) demonstrated that screening requires highly sensitive tests, whereas the next step—the confirmation of a diagnostic hypothesis that resulted from screening tests—requires high specificity. These authors further pointed out that few tests are equally well suited to both screening a population for a certain condition and confirming a positive result of a screening test.

The specificity of the phallometric test of pedophilia and homosexual hebephilia as a confirmatory test appears to be satisfactory. However, with offenders who do not admit to a corresponding erotic preference, the test does not appear to be very sensitive. This implies that a test outcome of gynephilia has much less predictive value than a test result of pedophilia or hebephilia.

The accuracy of a test—the likelihood that a positive outcome predicts the presence of the trait, disturbance, or disease that the test is designed to diagnose—does not, however, depend solely on the test itself, but also on the prevalence of this condition in the tested population. Use of a testing procedure for diagnostic reasons with (nonincest and nonexhibitionist) offenders against two or more children, in which a very high proportion of individuals can be supposed to have the trait or disorder, does not appear to be particularly useful. The test result cannot add very much to the probability that the trait or disorder is present. However, the usefulness of a testing procedure increases the nearer that the expected prevalence in the population approaches 50%. Among the results of the current study, this proportion is most closely approached by the pedophiles among offenders against only one female child. Therefore, systematic investigations of possibilities to increase the test's sensitivity will have to be geared particularly to this offender group.

## References

- Abel, G., Becker, J., Cunningham-Rathner, J., Mittelman, M., & Rouleau, J. (1988). Multiple paraphilic diagnoses among sex offenders. *Bulletin of the American Academy of Psychiatry and the Law*, 16, 153–168.
- Abel, G., Becker, J., Murphy, W., & Flanagan, B. (1981). Identifying dangerous child molesters. In R. B. Stewart (Ed.), *Violent behavior*.

- Social learning approaches to prediction, management, and treatment* (pp. 116–137). New York: Brunner/Mazel.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev). Washington, DC: Author.
- Freund, K. (1965). *Erotic Preferences Examination Scheme*. Unpublished manuscript.
- Freund, K. (1967). Diagnosing homo- or heterosexuality and erotic age-preference by means of a psychophysiological test. *Behaviour Research and Therapy*, 5, 209–228.
- Freund, K., & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology*, 57, 1–6.
- Freund, K., Diamant, J., & Pinkava, V. (1958). On the validity and reliability of the phalloglethysmographic diagnosis of some sexual deviations. *Review of Czechoslovak Medicine*, 4, 145–151.
- Freund, K., McKnight, C. K., Cibiri, S., & Langevin, R. (1972). The female child as a surrogate object. *Archives of Sexual Behavior*, 2, 119–133.
- Freund, K., Sedlacek, F., & Knob, K. (1965). A simple transducer for mechanical plethysmography of the male genital. *Journal of the Experimental Analysis of Behavior*, 8, 169–170.
- Freund, K., & Watson, R. (1990). Mapping the boundaries of courtship disorder. *Journal of Sex Research*, 27, 589–606.
- Freund, K., Watson, R., & Rienzo, D. (1988). Signs of feigning in the phallometric test. *Behaviour Research and Therapy*, 26, 105–112.
- Freund, K., Watson, R., & Rienzo, D. (1989). Heterosexuality, homosexuality, and erotic age preference. *Journal of Sex Research*, 26, 107–117.
- Frenzel, R., & Lang, R. (1989). Identifying sexual preferences in intra-familial and extrafamilial child sexual abusers. *Annals of Sex Research*, 2, 255–275.
- Griner, P., & Glaser, R. (1982). Sounding boards: Misuse of laboratory test and diagnostic procedures. *New England Journal of Medicine*, 307, 1336–1339.
- Griner, P., Mayewski, R., Mushlin, A., & Greenland, P. (1981). Selection and interpretation of diagnostic tests and procedures. *Annals of Internal Medicine*, 94, 559–600.
- Hollingshead, A., & Redlich, F. (1958). *Social class and mental illness: A community study*. New York: Wiley.
- Marshall, W., Barbaree, H., & Butt, J. (1988). Sexual offenders against male children: Sexual preferences. *Behavior Research & Therapy*, 26, 383–391.
- Marshall, W., Barbaree, H., & Christophe, D. (1986). Sexual offenders against female children: Sexual preferences for age of victims and type of behaviour. *Canadian Journal of Behavioural Science*, 18, 424–439.
- 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.
- Quinsey, V., Chaplin, T., & Carrigan, W. (1979). Sexual preferences among incestuous and nonincestuous child molesters. *Behavior Therapy*, 10, 562–565.
- Quinsey, V., Steinman, C., Bergersen, S., & Holmes, T. (1975). Penile circumference, skin conductance, and ranking responses of child molesters and “normals” to sexual and nonsexual visual stimuli. *Behavior Therapy*, 6, 213–219.

Received June 4, 1990

Revision received September 27, 1990

Accepted October 12, 1990 ■

### Harvey Appointed Editor of *Contemporary Psychology*, 1992–1997

The Publications and Communications Board of the American Psychological Association announces the appointment of John H. Harvey, University of Iowa, as editor of *Contemporary Psychology* for a 6-year term beginning in 1992.

Publishers should note that books should not be sent to Harvey. *Publishers should continue to send two copies of books to be considered for review plus notices of publication to*

PsycINFO Services Department, APA  
Attn: *Contemporary Psychology* Processing  
1400 North Uhle Street  
Arlington, Virginia 22201

Please note that all reviews are written by invitation.