

Sexual Abuse: A Journal of Research and Treatment

<http://sax.sagepub.com>

Statistical Adequacy of the Abel Assessment for Interest in Paraphilias

Lane Fischer and Gillan Smith

SEX ABUSE 1999; 11; 195

DOI: 10.1177/107906329901100303

The online version of this article can be found at:

<http://sax.sagepub.com/cgi/content/abstract/11/3/195>

Published by:



<http://www.sagepublications.com>

On behalf of:



ATSA*

[Association for the Treatment of Sexual Abusers](#)

Additional services and information for *Sexual Abuse: A Journal of Research and Treatment* can be found at:

Email Alerts: <http://sax.sagepub.com/cgi/alerts>

Subscriptions: <http://sax.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://sax.sagepub.com/cgi/content/refs/11/3/195>

Statistical Adequacy of the Abel Assessment for Interest in Paraphilias

Lane Fischer^{1,2} and Gillan Smith^{1,3}

A review of the psychometric properties of the Abel Assessment for Interest in Paraphilias (AAIP) is presented. The data supporting the AAIP's history, purpose, data, reliability, validity, and norms are reported, with the conclusion that its use with adults is tenuous at best and its use with adolescents is as yet unsupported. Pragmatic concerns are noted, along with recommendations for further research.

KEY WORDS: Abel Assessment; experimental design; paraphilias; penile plethysmograph; sexual arousal; sexual offending.

INTRODUCTION

Weinrott (1996) reviewed four major approaches to the assessment of sexual offenders: clinical impression, self-report, phallometry, and the Abel Assessment for Interest in Paraphilias (AAIP). He pointed out the need for an independent review of the use of the AAIP with offenders. This article undertakes such a review. The purpose of this study is to examine the research supporting the technical adequacy of the AAIP as currently marketed. The technical adequacy of any psychometric device depends on its history, its purpose, and the types of data it creates, as well as its reliability, validity, and norms.

AAIP HISTORY

Singer (1984) identified three stages of sexual attraction in males: (a) increased visual attention to the object of attraction, (b) movement toward that

¹Brigham Young University.

²320-J MCKB, BYU, Provo, Utah 84602.

³Norton Sound Corp., P.O. Box 1693, Nome, Alaska 99762.

object, and (c) resulting penile engorgement. Plethysmography purports to assess sexual attraction at Singer's third stage. Marshall (1996) noted the need for another assessment strategy because phallometry involves ethical problems inherent in presenting anyone, especially adolescents, with deviant sexual stimuli. The AAIP aims to circumvent such problems by measuring visual attention to possible objects of attraction consonant with Singer's first stage.

The AAIP is a creative instrument that uses self-report of attraction to 22 categories of possible sexual stimuli, as well as a surreptitious measure of sustained attention to photographs of the stimuli. It also includes a questionnaire regarding sexual behavior. This instrument is conceptually based on the research of Rosenzweig (1942), in which viewing time of pictures was shown to be a measure of sexual interest. Other researchers have utilized, adapted, and refined Rosenzweig's original concept (Brown *et al.*, 1973; Harris *et al.*, 1996; Love *et al.*, 1976; Quinsey *et al.*, 1996; Quinsey *et al.*, 1993; Ware *et al.*, 1972; Wright and Adams, 1994; Zamanski, 1956).

AAIP PURPOSE

The purpose of any measurement device is to provide data for decision making. Screening tests discriminate between normal and deviant groups within a population to signal an abnormality. Diagnostic tests discriminate between subjects within a deviant group to indicate specific pathology. Prognostic tests assess the risk to be at large or probability of treatment failure. The AAIP has been described as fulfilling all three functions. Abel *et al.* (1994) concluded that the AAIP effectively discriminated normals from pedophiles who had molested prepubescent or pubescent males. Abel (1995a) characterized the AAIP as a diagnostic test to measure interest in paraphilias. Subsequently, Abel (1997a) explained the test to judges as a prognostic test which provides "an objective evaluation of dangerousness" (p. 5) and enables professionals to "detect sex crime risk—particularly of offenders who target children" (p. 1).

AAIP DATA

The nature of the data determines the inferences which can be made from them. Numerical data may be created in various scales of measurement—nominal, ordinal, interval, or ratio—and may be norm-referenced or ipsative. In certain circumstances, ipsative scores can indicate intraindividual variation and rank-order comparisons between individuals. Interpretation of intraindividual variation is enhanced if one is aware of the underlying mean and standard deviation from which the ipsative scale was created. In the absence of such information, ipsative scales are analogous to ordinal scales. In such cases, interpretations are limited to conclusions that a subject possesses more of one attribute than another attribute, but

it is impossible to state how much more he possesses of one than another or when differences between attributes are significant.

Comparisons between subjects using ipsative scores are also limited to rank-order comparisons. Such descriptions are possible only if subjects' ipsative scores have been developed with reference to the scores of other subjects. In the absence of such references, between-subject comparisons are impossible.

The AAIP presents data in the form of ipsative z -scores for viewing time (VT). AAIP ipsative z -scores are created as follows: subjects view 160 slides divided into 22 categories of possible sexual attractors. The VT for each slide is recorded in milliseconds. Particularly deviant scores (outliers) are removed from the analysis by a formula which Abel maintains is confidential (personal communication, April 8, 1997). The mean VT per category is calculated, and the mean and standard deviation of the category means are then computed. The ipsative z -score for each category is the difference between the category mean and the grand mean of categories divided by the standard deviation of the category means.

The AAIP does not report the underlying raw score category means and standard deviation of category means for each individual and no reference is made to other subjects' rank order patterns of responses. Clinicians cannot approximate the meaning of any elevation other than the highest and lowest z -scores. The top-ranked category represents the longest average VT, while the lowest-ranked category represents the shortest average VT. Without knowing the specific raw score means and standard deviation of means for any given subject, it is tenuous to interpret any category other than the highest and the lowest scores. A single ipsative elevation may appear to be significantly deviant from the mean but in reality may be minuscule. Likewise, differences between categories may appear to be large but actually be quite small.

There are two errors in the interpretation of AAIP ipsative scores, both of which involve treating such scores as though they were norm-referenced interval scores. First, the AAIP presents data using a bar chart, with each of the 22 categories having a separate bar emanating from the ordinate of the graph. Figure 1 is an abbreviated facsimile of how AAIP scores are presented in the report to clinicians.

This type of presentation provides an illusion of normative reference which is inadvertently misleading. As examples, consider the interpretation of the data in Fig. 1. The interpretive guide states, "In Client A's case, his physiologic responding to adolescent African-American females is high (0.01)" (Abel, 1995a, p. 4). The graph makes it appear that Client A's score is one of the highest scores because the bar emanates from the ordinate at $-3z$. A z -score of 0.01 is almost exactly at the mean of all the scores and cannot be considered a deviant score. Rather than being high, it is, in fact, an average score.

The AAIP interpretive guide further states, "Client A's arousal to exhibitionism (EAF), voyeurism (VAF), and frottage (FRAF) all involve adult female targets and since the client's physiologic responding was not exceptionally high to any of these content areas, they should not be of therapeutic concern" (Abel,

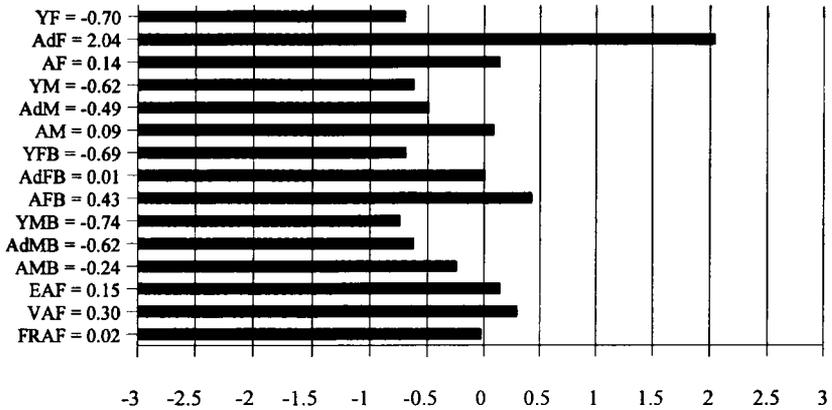


Fig. 1. Facsimile of AAIP graphic presentation of ipsative z-scores. YF, young Caucasian female (age 8–10); AdF, adolescent Caucasian female (age 14–17); AF, adult Caucasian female (over age 21); YM, young Caucasian male (age 8–10); AdM, adolescent Caucasian male (age 14–17); AM, adult Caucasian male (over age 21); YFB, young African-American female (age 8–10); AdFB, adolescent African-American female (age 14–17); AFB, adult African-American female (over age 21); YMB, young African-American male (age 8–10); AdMB, adolescent African-American male (age 14–17); AMB, adult African-American male (over age 21); EAF, exhibitionism toward adult Caucasian female; VAF, voyeurism toward adult Caucasian female; FRAF, frottage toward adult female.

1995a, p. 4). Rather than being “not exceptionally high,” the scores (EAF $z = 0.15$, VAF $z = 0.30$, FRAF $z = -0.02$) are very near the mean of zero. In fact, FRAF is slightly below the mean. These values are average but appear to be normative elevations because of the graphic presentation. The sum of ipsative z-scores always equals zero. Hence, the bar chart should emanate left and right (negative and positive) from a central zero point which represents the mean of z-scores. The correct presentation of this example is shown in Fig. 2. In this presentation it becomes apparent where the mean is and which scores actually deviate positively from it and which are truly below it.

The second error in interpreting ipsative scores is one of logic. The recommended interpretation of these scores is couched in the familiar logic of normative interval data. As an example, consider the interpretation of Client A's profile in the interpretive guide (Abel, 1995a). The z-score for the category AdF is 2.04, making it the top-ranked category. The interpretive guide states, “Adult males normally have slightly higher physiologic responding to adolescent females (AdF) than to adult females; the details of this arousal pattern can be found in Fierman . . . (1990). However the patient's arousal to adolescent females is exceedingly high” (p. 1). The terms “slightly higher” and “exceedingly high” are references to a normative interval when in fact the scores are ipsative elevations.

Abel (1995a) cited Fierman (1990) in supporting this use of the AAIP. However, Fierman did not present data to establish the normative expectation of scalar

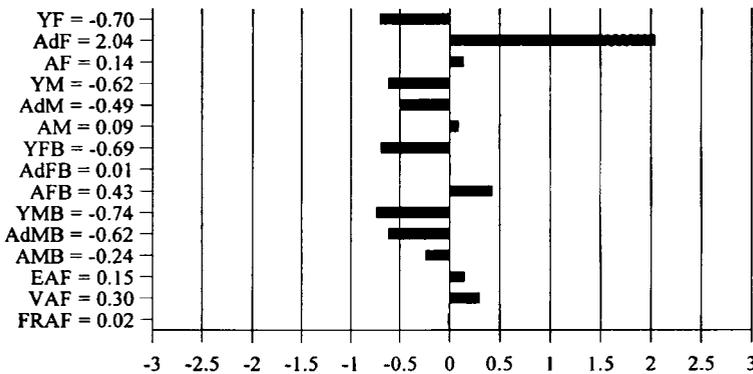


Fig. 2. The corrected graphic presentation of AAIP ipsative z-scores. YF, young Caucasian female (age 8–10); AdF, adolescent Caucasian female (age 14–17); AF, adult Caucasian female (over age 21); YM, young Caucasian male (age 8–10); AdM, adolescent Caucasian male (age 14–17); AM, adult Caucasian male (over age 21); YFB, young African-American female (age 8–10); AdFB, adolescent African-American female (age 14–17); AFB, adult African-American female (over age 21); YMB, young African-American male (age 8–10); AdMB, adolescent African-American male (age 14–17); AMB, adult African-American male (over age 21); EAF, exhibitionism toward adult Caucasian female; VAF, voyeurism toward adult Caucasian female; FRAF, frottage toward adult female.

differences between males’ attraction to adolescent females and their attraction to adult females. Fierman illustrated only that adult males were typically more attracted to adolescent females than to adult females. That is a statement of an ordinal relationship only. To interpret an elevation as “exceedingly high” requires interval data which neither Fierman nor the AAIP present.

Recently, Abel (1997b) devised a method to interpret the scores that first isolates the eight VT scores that are consistent with the client’s ethnicity only. Clinicians are to isolate the space between the lowest VT score and the highest adolescent or adult VT score. They are then to divide that space into thirds and consider any child category VT score that crosses the first third as a category that represents probable deviant sexual interest. Lacking a normative baseline, this arbitrarily establishes the starting point of interpretation at the lowest z-score value within a subset of eight categories within the profile. In the absence of a norm-referenced interval by which to judge category elevations, an arbitrary rule of thirds is imposed as a guide to interpretation. This represents a different strategy than has been used in the past, and while it would appear to allow clinicians to interpret ipsative scores as though they were norm-referenced, the essential nature of AAIP ipsative scores disallows such interpretations.

With regard to comparing clients with one another, Abel (1995b) stated that “since the raw reaction time measures vary from one client to another, if the raw values were used, it would be difficult to compare one client with another. . . . The

solution is to *z*-score the raw reaction time responding" (p. 53). That is correct only if the *z*-scores are based on norm-referenced means and standard deviations. Only then can clients be compared. The means and standard deviations used in AAIP ipsative *z*-scores vary between subjects. For example, if one client's mean VT is 2 sec and his standard deviation is 2 msec, while another client's mean is 20 sec with a standard deviation of 20 sec, the same *z*-score of 2 for both subjects would indicate something very different. If both clients had a *z*-score of 2 on a particular category, this would indicate an average VT of 2.004 sec for the first client (2-msec difference from the mean) and an average VT of 60 sec for the second client (40-sec difference from the mean). Thus, two quite different absolute behaviors are represented by the same ipsative *z*-score and would appear to be identical elevations from the means of their distributions. Because the underlying means and standard deviations are not reported and vary across subjects, there can be no comparison of one client with another.

This logical error can be corrected only with some difficulty. The ipsative scores need to represent a relative ranking of categories which is norm-referenced to the expected ranking of categories in a nonpathological population. Clemans (1956) articulated a strategy by which ipsative standard scores are created so that clients can be compared with one another and whereby a rough weighting of ranks can give a sense of deviation. If AAIP scores were transformed into ipsative standard scores, then a subject who goes through the material slowly could be compared with a subject who completes the task rapidly. The weighted ranking of categories could be compared. As it stands now, AAIP ipsative *z*-scores disallow such comparisons.

Another strategy would be to abandon ipsative scores in favor of norm-referenced *z*-scores. There may be a stable mean and standard deviation of VT in the normal population that would allow for comparison of subjects. This would require a very large stratified sampling of subjects across a wide geographic range. These subjects' scores would establish the normative mean, standard deviation, and ordinal position of each category. Subject VT's would then be captured in terms of normative *z*-scores that would allow interpretation and comparison across clients. The research to support this major shift has not been conducted. Currently, the presentation and interpretation of AAIP ipsative scores as though they were normative data could be misleading.

AAIP RELIABILITY

Reliability supersedes validity and sets the upper limit on any estimate of validity. If an instrument is not reliable, it cannot be valid. Reliability represents the consistency of measurement across conditions. Various types of reliability coefficients answer various questions about consistency of measurement across conditions.

Abel (1996, 1997a) and Abel *et al.* (1998) presented reliability data in the form of Cronbach's alpha for six categories: adult female = 0.86, adult male = 0.88 (ages 22+), adolescent female = 0.84, adolescent male = 0.89 (ages 14–17), young female = 0.90, and young male = 0.87 (ages 8–10). Abel (1995a) had earlier presented 21 categories. There are three problems with these reliability data. The first is that neither the 6 categories nor the 21 categories reported coincide with the 22 categories of the current test. There are categories missing and categories reported which have been removed from the test as currently marketed. It is unclear which slides have actually been used in the published test. Hence, the internal consistency data as reported do not apply to the AAIP as currently marketed.

A more compromising problem concerns the way the internal consistency coefficients were calculated in the first place. Abel *et al.* (1998) stated, "Visual reaction time data were screened for disparate responses and these were removed prior to data analysis" (p. 7). If one removes the disparate responses (outliers) before assessing the internal consistency of the categories, then the Cronbach's alphas are spuriously inflated.

The third problem is even more serious. Cronbach's alpha is a measure of internal consistency only. It measures whether items within a scale are homogeneous. The models in the AAIP slides are probably similar in gender and age within the reported categories. What Cronbach's alpha does not reveal is whether the clients tend to respond to the slides in a similar fashion time after time. If the test is being used to predict who has deviant sexual interests across time and setting, then test/retest reliability data are essential. Simply knowing that the slides within each category are similar does not answer the question of temporal stability of test results.

Abel (1996, 1997a) conducted a study to test whether subjects can falsify their responses to the AAIP. It is assumed that this was in response to concerns about the transparency of the instrument. This study might be misinterpreted as a validity study because the term "falsifying" is used in the title and it is placed in the validity section of the technical report, but it is actually a study of reliability. It represents an elegant study of the consistency of VT and self-report across several conditions using subjects who admitted or denied having abused children. The test was administered twice with different instructions to clients to either admit or conceal their attraction to the slides. The details of the analysis are sketchy in the report but seem to include a $2 \times 2 \times 2$ factorial ANOVA as well as correlations using a subset of ten categories of slides. The probable ANOVA variables include Type of Instruction (Conceal or Admit), Order of Instruction (Conceal then Admit or Admit then Conceal), and Type of Client (Admitters or Deniers). No data were presented from the ANOVA to justify the conclusion that, although subjects' raw VT changed across administration, the scores always changed in the same direction.

The standard for reliability is that coefficients should be greater than 0.80 (Anastasi, 1988). The average correlations of the two administrations for the

various categories ranged from 0.86 to 0.89 for self-report and from 0.56 to 0.74 for VT. These data indicate good reliability of self-report across conditions. The average coefficients for VT are not as robust. The coefficients for VT are averages which represent the central tendency of their distributions. No sample sizes, ranges, or variances were reported to enhance the interpretation of the average coefficients. These were not pure test-retest reliability, coefficients because the temporal variable was confounded by the instruction set and order of instruction. For the study to be a pure examination of test-retest reliability, the instructions would have to be identical in each administration.

Another problem with this analysis is that only 10 categories were used rather than the entire 22 categories of the test as currently marketed. It is undetermined which categories are stable.

AAIP VALIDITY

Validity confirms whether or not an instrument measures what it purports to measure and fulfills the purpose for which it is intended. Abel *et al.* (1994) reported a study that infers the AAIP can be used as a screening device. These investigators used six categories of slides: male and female children, adolescents, and adults. The models in the slides were partially clothed or nude. The authors reported that the slides differentiated normals from known child molesters. However, these are not the slides used in the AAIP as currently marketed. The current edition of the test uses only clothed models and presents 22 categories of attractors. Whether slides of clothed models can discriminate between deviant and normal groups, as did these six categories of nude models, is still unknown.

Abel and colleagues reported one other study of validity of the AAIP in separate documents in 1996, 1997a, and 1998. These data were presented in 1996 as evidence of validity for use with adolescents, in 1997a as validity data for use in the judicial system, and in 1998 as a comparison of the effectiveness of VT to plethysmography.

The data in the study were analyzed in two ways. The first analysis sought to determine whether an admitted child molester's AAIP results would correlate with his admitted target of abuse. The authors used only four categories of victims: females and males who were ages 14–17 or who were below age 14. It is unclear from the report whether the offenders were exclusive in their targets of sexual abuse or whether they preyed upon a range of categories of targets. It is equally unclear from the report how the AAIP scores identified a target victim. It is assumed that the highest category *z*-score identified that category as the target victim.

The study used a consistent subset of 42 slides from the various forms of a full set. There were seven slides each in six categories of possible sexual objects: females and males who were adults, ages 14–17, or below age 14. The study claimed to report the bivariate Pearson product-moment correlation coefficients

representing the association among the admitted victims, VTs, and self-report scores. In fact, no coefficients were reported.

This analysis, therefore, did not provide evidence of the validity of the 22 categories included in the AAIP. It did demonstrate that, using an unarticulated rule by which target victims were classified into four categories by a subset of 42 slides, there may have been a significant correlation with an admitted past target victim. However, no coefficients were presented to support even that conclusion.

The second consideration of these same data utilized discriminant analysis to create four prediction models to differentiate child molesters from other paraphilics. Abel (1996) stated that "the entire array of information contained in each screening instrument is utilized to 'predict' diagnosis" (p. 8). Abel *et al.* (1998) later used somewhat different language in reporting that the analyses "used all of the stimulus categories as independent variables" (p. 10). Discriminant analysis is a two-step procedure that first creates a multiple regression model in which variables are weighted according to their predictive power. Stepwise regression, which uses a mix of forward and backward selection of predictors from a pool, is used in this analysis. The resulting model is then applied to existing data with known cases to determine its actual effectiveness.

It is unclear from the report which variables were, or were not, available in the pool. The entire 22 categories could not have been used because only the 42 slides in the 6 categories were consistent in each administration. Furthermore, it is not clear which variables survived the stepwise regression procedures for each discriminant analysis. The models that produced the hit rates were not reported. Therefore, these data do not provide evidence of the validity of the AAIP.

AAIP NORMS

Norms establish the base of observations to which subsequent data are compared. A measurement device can be used ethically only with subjects who would reasonably have been found in the normative population. A problem may exist in using the AAIP with adolescents. The reliability and validity data are drawn largely from adults. Abel *et al.* (1998) reported that only 1.5% of the 157 participants were below age 18. As 1.5% of 157 is 2.35, evidently, only two participants were youthful offenders. No other studies have reported any adolescent samples. Even if the reliability and validity studies were perfectly sound for adults, they would not confirm how the test operates for adolescents.

CONCLUSIONS

We believe that the AAIP is a promising instrument based on an interesting concept. However, the evidence of its reliability and validity for use with adults is

weak as yet. There is no report of the technical adequacy of the test as currently marketed with 160 slides and 22 categories of sexual attractors.

The evidence of the reliability and validity of the AAIP for use with adolescents is still in question. Smith and Fischer (1999) have examined the test/retest reliability, as well as the screening and diagnostic validity, of the instrument for adolescents in residential and day treatment. A replication of that study is under way using incarcerated adolescents.

Abel (1997a) reported that approximately 300 therapists in 36 states and two foreign countries, as well as 8 states' judicial systems, currently use the AAIP. Users of the AAIP should be aware that there are possible dangers in using a nonvalidated instrument. The first is that judicial decisions based on these data are open to appeal and reversal. Similarly, treatment decisions based on these data may be compromised. Given the demands of managed care for data supporting accuracy of diagnosis and effectiveness of treatment, claims based on these data may be rejected by third-party payers.

Several recommendations spring from this review.

1. Research is necessary to establish a normative expectation which will make the data interpretable. Scores can be used as ipsatives if there is a normative expectation about what the ordinal patterns should look like regardless of means and standard deviations. With sufficient data, scores could be converted to true norm-referenced z-scores.
2. Internal consistency studies should be conducted on the full test as it is currently marketed.
3. Data should be gathered that justify the use of 22 categories as orthogonal constructs. If such data cannot be obtained, the test should be refined to match the true underlying structure of the categorical model.
4. Test/retest reliability studies should be conducted on the instrument as it is currently marketed.
5. Screening and diagnostic validity studies will also be necessary using the test as marketed.
6. All of the above analyses should be performed separately for populations of adults and adolescents.

The AAIP is a promising instrument based on a sound idea. The theory behind it is reasonable, the technology available, and the structure for further research in place. Further refinement is necessary so that its use may become a reliable and valid means to promote appropriate treatment of sexual offenders and possibly also a means to protect potential victims.

REFERENCES

- Abel, G. G. (1995a). *New technology: The Abel Assessment for Interest in Paraphilias*. Atlanta, GA: Abel Screening Inc.
- Abel, G. G. (1995b). *Abel screening system manual*. Atlanta, GA: Abel Screening Inc.

- Abel, G. G. (1996). *A new objective test for youthful offenders: The Abel Assessment for Interest in Paraphilias*. Atlanta, GA: Abel Screening Inc.
- Abel, G. G. (1997a). *Abel assessment for sexual interest: Judges' product information*. Atlanta, GA: Abel Screening Inc.
- Abel, G. G. (1997b). *Memorandum, 11/12/97*. Atlanta, GA: Abel Screening Inc.
- Abel, G. G., Lawry, S. S., Karlstrom, E., Osborn, C., & Gillespie, C. F. (1994). Screening tests for pedophilia. *Criminal Justice and Behavior, 21*, 115–131.
- Abel, G. G., Huffman, J., Warberg, B. W., & Holland, C. L. (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: A Journal of Research and Treatment, 10*, 81–85.
- Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.
- Brown, M., Amoroso, D. M., Ware, E. E., Pruesse, M., & Pilkey, D. W. (1973). Factors affecting viewing time of pornography. *Journal of Social Psychology, 90*, 125–135.
- Clemans, W. V. (1956). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs, 14*.
- Feierman, J. R. (Ed). (1990). *Pedophilia: Biosocial dimensions*. New York: Springer-Verlag.
- Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34*, 389–394.
- Love, R. E., Sloan, L. R., & Schmidt, M. J. (1976). Viewing pornography and sex guilt: The priggish, the prudent, and the profligate. *Journal of Consulting and Clinical Psychology, 44*, 624–629.
- Marshall, W. L. (1996). Assessment, treatment, and theorizing about sex offenders: Developments during the past twenty years and future directions. *Criminal Justice and Behavior, 23*, 162–199.
- Quinsey, V. L., Rice, M. E., Harris, G. T., & Reid, K. S. (1993). Conceptual and measurement issues in the phylogenetic and ontogenetic development of sexual age preferences in males. In H. E. Barbaree, W. L. Marshall, & S. M. Hudson, (Eds.), *The juvenile sex offender* (pp. 143–163). New York: Guilford.
- Quinsey, V. L., Earls, C., Ketssetzis, M., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology, 17*, 341–354.
- Rosenweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine, 4*, 150–158.
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *The Journal of Sex Research, 20*, 230–240.
- Smith, G. S., & Fischer, L. (1999). Assessment of juvenile sexual offenders: Reliability and validity of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11*, 207–216.
- Ware, E. E., Brown, M., Amoroso, D. M., Pilkey, D. W., & Pruesse, M. (1972). The semantic meaning of pornographic stimuli for college males. *Canadian Journal of Behavioural Science, 4*, 204–209.
- Wienrott, M. R. (1996). *Juvenile sexual aggression: A critical review*. Boulder, CO: Center for the Study and Prevention of Violence.
- Wright, L. W., & Adams, H. E. (1994). Assessment and sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment, 16*, 221–231.
- Zamasky, H. S. (1956). A technique for measuring homosexual tendencies. *Journal of Personality, 24*, 436–448.