

2007-12-11

Relationship between Key Variables in Penile Plethysmograph and Viewing Time Measures of Sexual Arousal in Sex Offending Adult Males

Lisa Loewinger Cloyd
University of Miami, lisa@wasatchonline.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

Cloyd, Lisa Loewinger, "Relationship between Key Variables in Penile Plethysmograph and Viewing Time Measures of Sexual Arousal in Sex Offending Adult Males" (2007). *Open Access Dissertations*. 58.
https://scholarlyrepository.miami.edu/oa_dissertations/58

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

RELATIONSHIP BETWEEN KEY VARIABLES IN PENILE PLETHYSMOGRAPH
AND VIEWING TIME MEASURES OF SEXUAL AROUSAL IN SEX OFFENDING
ADULT MALES

By

Lisa Loewinger Cloyd

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2007

©2007
Lisa Loewinger Cloyd
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

RELATIONSHIP BETWEEN KEY VARIABLES IN PENILE PLETHYSMOGRAPH
AND VIEWING TIME MEASURES OF SEXUAL AROUSAL IN SEX OFFENDING
ADULT MALES

Lisa Loewinger Cloyd

Approved:

Dr. Kent Burnett
Associate Professor
Department of Educational and Psychological
Studies

Dr. Terri A. Scandura
Dean of the Graduate School

Dr. Robert McMahon
Professor
Department of Educational and Psychological
Studies

Dr. Ray W. Winters
Professor of Psychology

Dr. Lane Fischer
Professor of Counseling Psychology
Brigham Young University

CLOYD, LISA LOEWINGER

(Ph.D., Counseling Psychology)

Relationship between Key Variables in Penile
Plethysmograph and Viewing Time Measures of
Sexual Arousal in Sex Offending Adult Males

(December 2007)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Associate Professor Kent Burnett.

No. of pages in text. (97)

Penile plethysmography is amongst the best measures for prediction of sexual recidivism and has been extensively researched. However, there are a variety of criticisms regarding use of penile plethysmography (PPG), including financial investment, significant training needs, lack of standardization, intrusive nature of the measure, extensive time for testing, and inability to use PPG with all groups of sexual offenders (e.g., women and adolescents). Viewing time has been studied, though not as extensively as PPG, regarding detection of sexual interest. This study examined the relationship between Monarch 21 PPG and Affinity viewing time variables, including responses to individual stimuli, a sex deviance differential, and a sex deviance ratio. It was predicted that there would be a significant positive correlation between key variables of the Monarch 21 PPG and Affinity. Overall, the associations found between the Monarch 21 PPG and the Affinity were significant, although the magnitudes of the associations were modest. When considering sexual deviance ratios and differentials (as is typically utilized in PPG literature), there was a small, but significant association between the Monarch 21 PPG and the Affinity. Consequently, it is recommended that more research be done comparing these two sets of measures, further evaluating the

Affinity and its' predictive validity, and examining the complex arousal pattern shapes, rather than focusing only on a series of discrete variables.

TABLE OF CONTENTS

| | Page |
|---|------|
| LIST OF TABLES..... | v |
| Chapter | |
| 1 INTRODUCTION | 1 |
| 2 REVIEW AND RELATED LITERATURE | 7 |
| Discussion of Penile Plethysmography..... | 7 |
| Early Studies of PPG | 7 |
| Modern Studies of PPG | 16 |
| PPG as a Predictor of Recidivism..... | 25 |
| Criticisms of Penile Plethysmography..... | 28 |
| Monarch 21 Addresses Validity Concerns | 36 |
| Discussion of Viewing Time | 38 |
| Early Studies of Viewing Time..... | 39 |
| Modern Studies of Viewing Time..... | 42 |
| Criticisms of the Abel Family of Viewing Time Measures..... | 54 |
| Introduction to Affinity Viewing Time Measure..... | 57 |
| Affinity Addresses Validity Concerns | 59 |
| Association between PPG and Viewing Time Measures..... | 59 |
| Improvements over Letourneau Study..... | 62 |
| Statement of the Problem..... | 63 |
| Research Questions..... | 65 |
| Hypotheses..... | 65 |
| 3 METHOD | 71 |
| Design | 71 |
| Participants..... | 71 |
| Measures | 71 |
| PPG | 71 |
| Viewing Time Measure..... | 73 |
| Procedure | 75 |
| 4 RESULTS | 77 |
| Analyses of Hypotheses..... | 78 |
| Hypothesis One..... | 78 |
| Hypothesis Two | 79 |
| Hypothesis Three | 80 |
| Hypothesis Four..... | 85 |
| Hypothesis Five | 86 |

| | | |
|---|--------------------------------|----|
| 5 | DISCUSSION..... | 87 |
| | Hypothesis One..... | 87 |
| | Hypothesis Two..... | 88 |
| | Hypothesis Three..... | 89 |
| | Hypotheses Four and Five..... | 91 |
| | Strengths and Limitations..... | 92 |
| | Future Research..... | 93 |
| | References..... | 94 |

LIST OF TABLES

| | Page |
|---------------|------|
| Table 1 | 66 |
| Table 2 | 67 |
| Table 3 | 68 |
| Table 4 | 69 |
| Table 5 | 79 |
| Table 6 | 82 |
| Table 7 | 83 |
| Table 8 | 84 |

CHAPTER I

Introduction

Several recent cases of recidivism in convicted sexual offenders have garnered national attention (Jessica Lunsford, Sarah Lunde, Dylan and Shasta Groene) and contributed to public outcry. Consequently, there has been much recent federal legislation to draft and enact new laws to protect children from those who sexually offend, including federal legislation such as the Children's Safety and Violent Crimes Reduction Act of 2005 (H.R. 4472), the Sex Offender Registration and Notification Act (H.R. 3132), Dru Sjodin National Sex Offender Public Database (S.792), and the Jacob Wetterling, Megan Nicole Kanka, and Pam Lychner Sex Offender Registration and Notification Act (S.1086). States are also drafting and enacting new laws such as "Jessica's Law" (currently in 18 states) and laws requiring sex offenders live a minimum distance (from 1000 feet to 2500 feet) from a variety of areas where children may congregate (e.g., schools, daycares, parks). Preventing further victims of sexual abuse is clearly of concern to many citizens and lawmakers.

In order to prevent further victims and ensure the best interests of the public in general, decisions must be made by psychologists, judges, juries, and probation officers regarding treatment and placement of those who have sexually offended. In order to make these critical decisions, however, these individuals need data regarding potential risks, diagnoses, and prognostic indicators.

Penile plethysmography (PPG) is the standard method utilized in the assessment and treatment of adult males who have offended sexually (Marshall, 1996; Kalmus & Beech, 2005). PPG is used to assess a male's sexual arousal with regard to age and

gender. Research on PPG has been ongoing for over 40 years (i.e., since 1963) with a variety of groups in order to effectively discriminate sexual arousal patterns between different individuals. Research has shown that deviant sexual arousal, as measured by the PPG, is consistently amongst the top two predictors of sexual recidivism in adult males who have sexually offended (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). Consequently, it is critical that methods of measuring sexual deviancy be further refined to assist in prediction of recidivism in order to assist decision-makers with determining appropriate placement and treatment of those who have sexually offended.

Despite the research showing PPG is such a useful predictor of sexual recidivism, the method has several significant drawbacks. One significant drawback of PPG is a large financial investment, as PPG systems cost thousands of dollars. Additionally, proper administration of PPG requires many hours of training for each individual who administers the PPG. Each administration of the PPG may also require several hours of time of the client sitting in a chair while attached to the PPG system. Another problem with the PPG is the lack of standardization (Quinsey & Laws, 1990; Howes, 1995; Marshall & Fernandez, 2000). Researchers have also raised concerns regarding the ability of subjects to “fake” the results of the PPG (McAnulty & Adams, 1991; Kalmus & Beech, 2005). Finally, a significant concern regarding penile plethysmograph is the intrusive nature of the measure. That is, the client has a device directly attached to his penis, measuring changes in either circumference or volume, while being exposed to a variety of stimuli.

Viewing time is another method of measuring an individual’s sexual interest, which was initially researched in 1942 (Rosenzweig, 1942). An association has been

shown between the amount of time a subject views particular stimuli and their sexual interest in stimuli of a similar age and gender (Chaplin, Rice & Harris, 1995; Love, Sloan, & Schmidt, 1976). Additionally, viewing time measures have the advantage over PPG due to their ease of administration, little training required of administrators, lack of invasiveness, and less significant financial investment. If viewing time can be shown to be similarly effective in predicting risk of sexual recidivism as PPG, it may be a good replacement for PPG or it may add to the ability to predict recidivism if combined with PPG.

Although much research has shown viewing time is an effective measure of sexual interest, there have been a variety of criticisms regarding this research. First, most of the recent research regarding viewing time has been conducted in the lab of one developer of the measure (i.e., Gene G. Abel, M.D., who developed the Abel Assessment for Sexual Interest and Abel Assessment for Interest in Paraphilias). The data used in this research has not been made available to other researchers to be verified (Fischer and Smith, 1999). Indeed, in only one study was the original, “untrimmed” data provided to a researcher outside of Dr. Gene Abel’s lab (Letourneau, 2002). Additionally, this measure of viewing time has been shown by some researchers to have poor test-retest reliability (Kaufman, Rogers, & Daleiden, 1998; Smith & Fischer, 1999).

A recent development in measuring viewing time is the Affinity (Glasgow & Croxen, 2003). The most significant advantage of this measure is the ready accessibility of the data to researchers. Indeed, this instrument has been developed by researchers and the data remains with the initial administrators, rather than being sent to a central office for scoring and interpretation, as with the Abel Assessment for Sexual Interest. Affinity

stimuli include images of individuals of varying ages and both genders in typical daily situations. The raw data becomes part of a data set maintained by the researchers gathering the data. Consequently, all data remains available and data gathered via the Affinity may be examined and replicated by a variety of researchers.

Letourneau (2002) was the only researcher outside of Dr. Abel's lab who was provided original, "untrimmed" data from the measures she administered. As the PPG has been shown to be a good predictor of sexual recidivism, but has significant limitations (as previously discussed) and viewing time is an effective, but new, measure, Letourneau examined the concordance between the two measures. In this study, Letourneau administered both PPG and a viewing time measure to 57 adult males incarcerated in a maximum security military prison. The mean sentence for subjects was 9.47 years and the average length of incarceration at the time of testing was 3.37 years. Subjects were administered the penile plethysmograph while being exposed to 16 audio stimuli including male and female adults and children. The age of the children were not specified, but the subjects were instructed to "think of a child just the age that you like." The viewing time measure included visual stimuli consisting of adult men and women, and children in the following discrete age ranges: 2-4, 8-10, and 14-17. Letourneau then computed correlations between the results of both the PPG and viewing time measure. The researcher also created a 2x2 grid in which the identified sexual preference from each measure was compared to the victim choice, thus identifying the number of hits and misses with regard to classification. Letourneau found that the measures were significantly associated and have similar discriminative ability when considering

offenders against male children. However, the measure of viewing time was better able to discriminate those who offended against female children than was the PPG.

Despite the seminal nature of Letourneau's (2002) study, there were several significant flaws. First, her sample size was quite small, as she had only 57 subjects. Letourneau's study also had problems with generalizability in several areas. First, the subjects were all military personnel and were all incarcerated for their offenses. Many males who sexually offend are not incarcerated. Additionally, the nature of their occupation is highly specific and may not necessarily generalize to the overall population. Another concern with Letourneau's study is her use of the Abel Assessment for Sexual Interest, as the data obtained in this measure are not readily available to researchers. Consequently, verification and replication of the data is quite difficult. Finally, Letourneau utilized a measure of penile plethysmography that is not standardized, which is one of the significant critiques of the PPG overall.

There has been only one study of the concordance between PPG and viewing time measures of sexual arousal and interest (Letourneau, 2002). Unfortunately, this study utilized a small and specific sample, unstandardized PPG, and the Abel Assessment for Sexual Interest, which has significant documented problems. The purpose of the current study was to examine the concordance between a PPG system and viewing time using a larger and more diverse sample, standardized PPG system (i.e., Monarch 21), and a better measure of viewing time (Affinity).

The current study offered an improvement over Letourneau (2002) in several areas. First, subjects in the present study came from a general population, some of whom remained in outpatient treatment, some of whom went on to be incarcerated, and others

who have been released from prison or jail settings. Consequently, results of the current study were likely to be more generalizable than Letourneau's study. Another advantage of the current study was the number of subjects included. As previously noted, Letourneau's study included a total of 57 subjects; the current study had 96 subjects. Another advantage of the current study was the utilization of a PPG method that addresses issues related to "faking" which have been researched. A final advantage of the current study's PPG was utilization of a standardized method of penile plethysmography. That is, the core stimuli are consistent each time this PPG system is utilized, in comparison with another PPG system that is not packaged with any stimuli. Consequently, the current PPG system addresses several concerns researchers have expressed. Additionally, the current study utilized the Affinity viewing time measure so data gathered are readily available to the researchers.

The current study examined the relationship between the results of the Affinity (a measure of viewing time) and a standardized PPG system (the Monarch 21). It was hypothesized that the two measures would be significantly correlated in regard to key variables of sexual interest and sexual arousal by age groups and gender of the stimuli.

CHAPTER II

Literature Review

Discussion of Penile Plethysmography

Phallometry or penile plethysmography (PPG) has been the main method of measuring sexual arousal or deviance in sexual offenders for over 20 years (Marshall, 1996; Kalmus & Beech, 2005; O'Donohue & Letourneau, 1992). PPG is a method whereby penile tumescence is measured by placement of a gauge over the subject's penis while he is exposed to a variety of stimuli chosen by the evaluator. A man's sexual arousal to differing scenarios or stimuli is purported to be measured by measuring penile tumescence. An underlying assumption of PPG research is that penile tumescence is associated with sexual preference, which may predict sexual behaviors.

Early Studies of PPG

The penile plethysmograph was developed by Kurt Freund (1963) to measure men's sexual arousal, specifically to discriminate between heterosexuals and homosexuals. Freund (1967a and 1967b) also began measuring sexual arousal of males who molest children in 1967. Males who had molested children viewed sixty slides with images of males and females classified as children, adolescents, and adults. Unfortunately, the specific age range of each category was not provided. Control heterosexual subjects who were presumed to have "normal" sexual arousal responded similarly to adult and adolescent females, which was significantly greater than arousal to female children ($p \leq 0.05$). However, males who offended against children evidenced an arousal pattern with most arousal to adult and child females, which was significantly greater than arousal to female adolescents ($p \leq 0.05$). Though these results are different

than common perception would predict (as there was similar arousal between adults and age-inappropriate adolescent females in control males), these results do evidence that there is a different overall arousal pattern which appears to discriminate between presumably “normal” and pedophilic arousal. That is, “normal” subjects show significantly more sexual arousal to adults and adolescents than children, whereas males who offended against children show significantly more arousal to adults and children than adolescent females. Consequently, it may be beneficial to study key variables of sexual arousal and interest, in contrast to simply examining differences in arousal levels to different age groups, as (in this example) both groups showed significant arousal to adults, whereas they differed in their responses to females of other age categories.

Subsequently, many other researchers have compared PPG responses between sex offenders and those who have not offended sexually, as well as between different types of sex offenders. Though results generally reveal differing arousal patterns, results have been varied between studies.

Freund and Blanchard (1989) studied PPG differences between men who offend against children and those who offend against adult females. The researchers excluded men who had offended against both adults and children. Stimuli consisted of nine categories: adult women, young adolescent females (no Tanner stage specified), 8 to 11-year-old girls, 5 to 8-year-old-girls, adult men, young adolescent males (no Tanner stage specified), 8 to 11-year-old boys, 5 to 8-year old boys, and landscapes (neutral stimuli). Men who did not respond a predetermined amount were excluded from the analysis as non-responders. Specificity was found to be 95% though sensitivity was estimated at 55%. The researchers hypothesized that 55% is truly the lowest possible sensitivity, as

individuals who were diagnosed by the measure may truly not have been pedophilic, but offended against children for other reasons, such as lack of an age-appropriate partner. The efficiency of non-admitter classification was 69%, representing an improvement over chance in this group that is of most interest.

Freund and Watson (1991) performed a follow up study to the Freund and Blanchard (1989) study, which compared PPG results between males who offended against children and males who offended against only adult females. The researchers hypothesized that men who offend against multiple child victims are more likely to be truly pedophilic than those who offend against only one child, which artificially decreases the sensitivity of detection of pedophilia. Consequently, they argued that division of offenders into groups with differing numbers of victims could provide better estimates of sensitivity. The research showed sensitivity of 45% in those who offended against a single female child, 75% in those who offended against at least one female child and one female pubescent, and 87% in those who offended against a single male child. Specificity was low for control subjects who had not been accused of sexually offending behavior (80.6%), but was 96.9% for subjects who had offended against only adult females.

The authors argue this study suggests that discriminative PPG control subjects should be individuals who sexually offended against adults, in comparison with “normal” subjects who have not been known to offend. However, this recommendation appears misguided, as this study highlighted similarity between individuals with presumably normal arousal. Instead, it is even more critical to understand normal arousal so that better discrimination is provided between those who offend against children and normal

controls. That is, it is still unknown how “normal” individuals would have responded and the differences between individuals who offended against children, against adults, and those who are not known to have offended. This study also did not provide efficiency estimates, raising questions regarding PPG’s predictive improvement over chance. Nonetheless, this study does support the notion that PPG may validly discriminate between those who molest children and control groups, though the degree of validity remains unclear.

Marshall, Barbaree, and Christophe (1986) conducted a study in which they compared arousal of individuals who have offended against unrelated female children, related female children, and a nonoffending community sample of controls. Subjects in both offender groups either admitted guilt for their offenses or there was “incontrovertible evidence” of their offense and they were subsequently convicted for their offense. There were a total of 61 subjects, but 22% of those who offended against unrelated children and 34% of those who offended against related children were excluded from the analysis due to low responding. The decision to exclude “nonresponders” or “low responders” from studies occurs frequently, due to concerns related to the reliability and validity of the data. However, it may be argued that these are the offenders who most need to be studied due to their low responsiveness and the possibility of this pattern of responding being due to an attempt to falsify assessment results.

In the aforementioned study, testing data was gathered over three sessions (Marshall, Barbaree, and Christophe, 1986). The first session involved completion of Raven’s Progressive Matrices, collection of demographic information, and consent. During the second session, subjects were shown slides of nude females ranging from

three to 24 years of age with each stimulus shown for two minutes. Additionally, the researchers noted images of 12 and 14 year olds evidenced immature secondary sexual characteristics, while slides of females 16 and older evidenced mature secondary sexual characteristics. Though the stimuli were reported based upon age, it was important to describe the visual maturity of the individuals as it suggests homogeneity of maturity rather than age, which could be different than suggested by appearance. During the third session, subjects listened to six audiotapes of a male speaking in second person discussing sexual contact with a female child. Three segments involved agreement to engage in sexual contact which included fondling in one tape, mutual oral-genital contact in another tape, and sexual intercourse in the third tape. The remaining three audiotapes all involved sexual intercourse and aggressive interactions in which the child refused, physically resisted, and informed the narrator she was being hurt. The degrees of violence varied in the tapes, as one involved threats of physical violence, another involved physical force, and the third included slapping and punching. After exposure to the stimuli was complete, subjects were shown a sexually provocative videotape and were asked to produce a full erection. Consequently, each subject's response during testing was able to be expressed as a percentage of full erection.

Review of overall patterns of the three groups in the aforementioned study revealed differences amongst each group (Marshall, Barbaree, and Christophe, 1986). The control group showed most arousal to adults and little arousal to children. Those who offended against unrelated children evidenced most arousal to children and adults, but little to adolescents. The subjects who had offended against related children showed little arousal to children, but their arousal gradually increased as age increased, as compared to

the sharp increase at 16 years evidenced by the control group. It appears those who offended against related children did not differentiate between children to the degree that the control group did. Experimenters then created deviance quotients in which arousal to deviant stimuli (children and adolescents) was expressed as a ratio to nondeviant (adult) stimuli. Utilization of this procedure was better than chance, but did not appear to be as useful as examining key variables that may be associated with a general pattern (such as deviance differential, deviance ratio, or the shape of the arousal curve). Instead, it is noteworthy that it was not a particular data point that differentiated between the groups, but their overall differences in responsiveness to stimuli of representing different age groups. This observation supports the notion that understanding a male's sexual arousal may be better understood by the aforementioned key variables, rather than simply arousal to children, adolescents, or adults.

Quinsey and Chaplin (1988) compared sexual arousal of 25 offending subjects with 14 controls who were not known to sexually offend. The offending subjects were inpatients in a mental health hospital, 23 of whom were diagnosed with personality disorders, one of whom was psychotic, and the diagnosis of the final subject was unknown. The comparison group consisted of eight males recruited from the community and six inpatient males from the same institution, two of whom were diagnosed with personality disorders and four of whom were psychotic. The inclusion of psychotic individuals in this study is a concerning confound, as it is unclear if the subjects actually responded to the stimuli or if their psychosis was interfering with their processing of the stimuli.

Subjects were administered three “warmup” stimuli (two nonsexual heterosexual interactions and one scenario of consenting adult heterosexual activity) to acclimate to the experimental situation (Quinsey and Chaplin, 1988). Stimuli of interest consisted of scenarios with matched gender for sexual orientation involving different levels of interaction with an eight or 10-year-old consisting of sexual intercourse with no violence and passive resistance, coercive intercourse utilizing threats and physical force, violent sex with needless physical injury, and nonsexual violence. Researchers recorded the maximum response and subtracted the baseline at the beginning of each segment to establish a change score for each segment.

Results of the Quinsey and Chaplin (1988) study showed there are differences in PPG responses between the offender and control groups. The control group responded exclusively to consenting adult stimuli, whereas those who offended against children responded significantly more to children. Indeed, child molesters responded significantly more to the agreeing child stimuli than they did to consenting adult stimuli. Interestingly, there was no significant difference between experimental subjects’ response to the agreeing child stimuli and the violent sexual interaction with children, though they responded significantly more to the agreeing child stimuli than to the nonsexual violent stimuli. Additionally, the more violent the subjects were historically, the greater their average response was to the coercive stimuli, though not the violent stimuli. Results of this study suggest subjects who offended against children were not discriminating between stimuli, as they were generally aroused to all child sexual scenarios as well as adults (though the offenders were less aroused than the control subjects to adult stimuli),

further highlighting the need to examine the particular variables reflective of a pattern of responding, rather than simply individual data points.

Barbaree and Marshall (1989) performed a study comparing those who offended against related female children (incest offenders), those who offended against unrelated female children (child molesters), and a control group of subjects who were not known to offend sexually. Subjects were excluded from the study if they did not achieve 10% of full erection based upon self-report of full erection to a sexually provocative videotape presented at the conclusion of the study. Stimuli consisted of nude images of females from three to 11 years (“child”), 12 to 14 years (“pubescent”) and 16 to 24 years (“adult”). Researchers blindly sorted the results by the profile shape and determined there were five identifiable shapes: 1) adult, large response to adults and minimal response to all other stimuli; 2) teen-adult, large response to adults and pubescents with moderate arousal to older children; 3) non-discriminating, similar arousal to all age groups; 4) child-adult, substantial responses to adults and child, but none to pubescents and 5) child, only significant response to children.

In Barbaree and Marshall’s study (1989), more than one in three subjects from the child molester group evidenced the child arousal pattern and no one from any other group evidenced this pattern. This child pattern may represent a subset of individuals who offend who are true pedophiles and offend exclusively due to their arousal, rather than other psychological motivators. Additionally, the non-discriminating group may have a problem with disinhibition, as they are equally aroused to all individuals. Another possible explanation for the non-discriminating pattern is that these individuals were able to suppress their sexual arousal to children, though they could not enhance their arousal

to adults. When considering those who evidenced the child-adult pattern, it is noteworthy that the child responses were slightly less than the adult responses and none of the control subjects evidenced this pattern. However, if the researchers had used a simple difference score between the child and adult arousal, these subjects would have appeared more sexually interested in adults. Consequently, for this study, it does not appear that a simple comparison of specific data points would have yielded as much discrimination between groups as did analysis of the general response pattern. Indeed, it appears more helpful to consider data points in context with each other, rather than exclusively as the sole indicator of sexual arousal. Results of this study suggest that PPG may be used to differentiate between some groups of sexual offenders, as the arousal pattern was significantly different. However, some child molesters and 40% of the incest offenders evidenced the adult arousal pattern, suggesting this measure may not be used alone to determine group affiliation.

Utilizing a volumetric of measurement, Frenzel and Lang (1989) conducted a large study of 191 subjects including those who reported heterosexual orientation who offended against unrelated children (heterosexual child molesters), those who reported heterosexual orientation and offended against related children (heterosexual incest offenders), those who reported homosexual orientation and offended against unrelated children (homosexual child molesters), and controls not known to offend. Stimuli included males and females between six and fifteen years of age and adults aged 18 to 25 years.

Results of the Frenzel and Lang (1989) study supported the notion that there are differences in PPG responses between the aforementioned groups. The control group

showed the lowest responsiveness, the homosexual child molester and heterosexual incest offenders showed moderate arousal, and the heterosexual child molester group showed the most responsiveness in terms of the absolute arousal levels. For the control group, responses to all males and the youngest females (aged six to eight) were not significantly different than their responses to neutral stimuli; as the age of females increased, so did the control group responsiveness. Heterosexual child molesters did not respond significantly differently to males than to neutral stimuli and their responses increased as the age of the females increased. However, heterosexual child molesters' responses to adolescent females were not significantly different than their responses to adult females, suggesting they may not be discriminating between age-inappropriate adolescents to the extent that "normal" subjects do. Heterosexual incest offenders evidenced a similar pattern to control subjects. Finally, homosexual child molesters were most aroused to 13 to 15-year-old boys, followed by nine to 11-year-old boys; interestingly, their responses to all females were higher than their responses to boys between six and eight years, adult males, and neutral stimuli. Finally, it is noteworthy that only 68% of the control subjects showed their largest single response to adult females. Consequently, this study could not discriminate well based simply upon the single highest response. Instead, the overall arousal pattern appears to better demonstrate differences between groups, though this was not thoroughly addressed in this study.

Modern Studies of PPG

Harris, Rice, Quinsey, Chaplin, and Earls (1992) studied differences between arousal of those who have offended and control subjects. Researchers used three groups of data sets. The first data set consisted of three groups: homosexual child molesters,

heterosexual child molesters, and control subjects. Gender preference was determined by the gender of the latest victim, if there were multiple victims. Half of the subjects were inpatients at a psychiatric hospital and half were outpatients. Stimuli consisted of slides of males and females ranging from one to 17 years old, as well as adults and neutral images. The second data set consisted of child molesters in an inpatient hospital and controls. They were presented with audiotapes with neutral interactions, consenting adult sexual interactions, nonsexual violence against a child, and sexual activities with a child varying in interaction between passive, coercive, and violent. The third data set was the largest sample and included 77 rapists, 29 non-sex offenders, and 15 control subjects. Stimuli were audiotaped stories of heterosexual nonsexual interaction, consenting adult sexual interaction, stories involving rape, and nonsexual beatings.

Harris et al. (1992) utilized a deviance differential and a deviance ratio, in which the deviant categories were compared to the nondeviant or consensual categories. In all three data sets, offenders evidenced significantly more arousal to deviant than to nondeviant stimuli and control groups evidenced their highest arousal to nondeviant stimuli. However, in the third data set involving rapists, control subjects had larger absolute responses overall, thus evidencing more responsiveness to deviant stimuli than did the offenders. Nonetheless, the pattern of responses of control subjects continued to evidence significantly more arousal to nondeviant than to deviant stimuli. These results highlight the importance of viewing responses in context, rather than considering only the absolute level of response or simply comparing data points independently. Additionally, researchers removed nonresponder data from the data set to determine if the group discrimination was improved. Interestingly, the researchers concluded that the exclusion

of these subjects did not improve the validity even when only 30% of the subjects remained in the study, noting, “phallometric profiles of low responders are just as valid as those of relatively high responders.” Finally, discrimination between groups was not improved by inclusion of neutral stimuli. Instead, more than 80% of subjects who responded to neutral stimuli were from the offender group, suggesting responsiveness to neutral stimuli may suggest subjects are faking responses.

Chaplin, Rice, and Harris (1995) compared sexual responding of 15 males who offended against children, both related and unrelated, with 15 control subjects in order to examine sensitivity and specificity of their stimuli. The stimuli the researchers used are quite different than typical stimuli. Stimuli consisted of 22 audiotapes involving different narrators and different scenarios and two warmup audiotapes. Two scenarios involved nonsexual interactions, one narrated by a girl and one by a man. Four stimuli involved sexual interactions between a female child and an adult male, narrated from his perspective and no evidence of her resistance, which is typical of PPG stimuli. Four stimuli also involved consenting adult heterosexual interaction, narrated from the female’s perspective. The stimuli that were most unusual were the 12 stimuli narrated by the female child describing sexual interaction with an adult male. Four of the aforementioned scenarios described a passive and nontraumatic sexual interaction, four involved a moderately traumatic sexual encounter in which the child complains, is fearful, and uncomfortable, and the final four depicted an extremely traumatic experience in which the child experiences significant pain and “is terrorized.” Controls showed significant differences between their responses to consenting adult sex and all other stimuli, whereas child molesters showed no significant difference between any sexual

categories. In terms of absolute arousal values, child molesters showed significantly more arousal to all deviant stimuli than did controls, but there was no significant between-group differences in arousal to adults.

Researchers used a deviance differential to examine the difference between the groups, in which the highest deviant response was subtracted from the highest nondeviant response, and found only one person misclassified, resulting in 93% sensitivity and 100% specificity. When the two highest responses to the deviant stimuli were averaged and then subtracted from the highest response to nondeviant stimuli, researchers achieved a sensitivity and specificity of 100%. However, it is noteworthy that researchers chose an optimal cut point based on their knowledge of a 50% base rate. Though averaging the two highest deviant responses somewhat lowers the magnitude of the response (which typically included the passive child and moderate or extreme trauma stimuli), it is noteworthy this method was better able to differentiate between the groups. It is possible this increased sensitivity may not be a reflection of the arousal to the child stimuli overall, but is instead a lack of inhibition of arousal to traumatic stimuli. That is, control subjects' arousal was significantly inhibited by the traumatic nature of stimuli, which did not occur for child molesters. This lack of differentiation may contribute to the tendency to sexually offend. Additionally, it is again noteworthy that child molesters did not appear to differentiate between sexual stimuli, as compared to normals, also raising the possibility child molesters were unable to enhance their responses to adults, though they were able to inhibit their responses to children. Overall, however, results of this study suggest that, given particular stimuli, PPG is extremely useful in differentiating between those who offend and normal controls.

Barsetti, Earls, Lalumiere and Belanger (1998) compared arousal of males who offended against family members (incest offenders), males who offended against unrelated children (though also possibly against related children) referred to as child molesters, and control subjects. Subjects were excluded who did not have any higher response than to neutral stimuli. Stimuli were audiotaped translations of stimuli used in Quinsey and Chaplin (1988) with an additional four stimuli depicting incestuous sexual contact with a female or male child. Raw scores were converted to z scores for analysis. There were no differences between the groups based upon their overall responsiveness, though their arousal patterns were different. Control subjects responded most to consensual adult sexual interactions and no further significant differences between neutral and other stimuli were found. Incest offenders and child molesters evidenced significantly more arousal to the following categories than to neutral stimuli: consenting female adult, passive child, incestuous contact with a passive child, and sexual coercion of a child. Additionally, child molesters also responded significantly to the rape of a child. Controls also evidenced significantly more arousal to consenting adult stimuli than did incest offenders and child molesters, but no differences were found between incest offenders and child molesters. Those who offended against children appeared to show a nondifferentiating arousal pattern between groups, rather than increased arousal to children relative to adults, further supporting the need to examine the results in context of each other, rather than only particular data points.

Grossman, Cavanaugh, and Haywood (1992) compared arousal of 31 outpatient males accused of offending against either unrelated children or related child and 21 control subjects not known to offend sexually. Subjects were shown slides of nude and

partially clothed adult females, and male and female children between four and 12 years of age. Researchers used percent of full erection as a measure of responding, utilizing a 30mm circumferential increase as full erection. Slides were divided simply into adult versus child groups. Controls evidenced little arousal to children but significant arousal to adults. Child molesters showed more than two times the arousal to child slides than did the control subjects, but also evidenced significantly less arousal to adults than did normal controls. Indeed, it is noteworthy that the child molesters had a generally flat arousal pattern, in which they did not significantly differentiate between child and adult stimuli. Incest offenders evidenced similar arousal to children as did controls, but their arousal to adults was less than half the arousal of the controls to adults. Indeed, the incest offenders' arousal pattern was quite similar to the arousal pattern of child molesters, though they responded less overall than child molesters did. Thus, in terms of absolute arousal to children, incest offenders were no different than control subjects. However, their arousal to adults was significantly less than the normal controls' arousal to adults. Results of this study further support the notion it is critical to examine the overall responses in context with each other rather than simply examining independent data points.

PPG has also been shown to discriminate between subjects who offended against girls and boys, classified in this study as boy-object pedophiles and girl-object pedophiles (Laws, Hanson, Osborn, & Greenbaum, 2000). In this study, there were 124 initial subjects engaged in an outpatient treatment program, but 52 were excluded due to being classified as bisexual, not completing all three components of the study, or attrition from treatment. Measures included a card-sort task, PPG with slide stimuli, and PPG with

auditory stimuli. Visual PPG stimuli included three slides for each category of the six Tanner stages of males and females; auditory stimuli depicted fondling a child, non-protesting sexual intercourse with a child, coercive intercourse with a child, rape of a child, sadistic sexual assault of a child, physical (nonsexual) assault of a child, and consenting adult intercourse. The final score for a given stimulus was the maximum arousal during the stimulus presentation minus the minimum measure during the same stimulus. Additionally, researchers utilized “relative arousal,” which was the ratio of deviant arousal to nondeviant (consenting) arousal.

Results of this study suggest that combining all three measures produces the most effective method of discriminating between groups, though there were some potential confounds with this study (Laws, Hanson, Osborn, & Greenbaum, 2000). When discriminating between the groups, inclusion of female children added no discrimination; however, when male children were initially considered, adding female child responses significantly added to correct classification. Utilizing all three measures (PPG with audio, PPG with video, and card-sort) produced the best classification, as 100% of the girl-object pedophiles and 70% of the boy-object pedophiles were correctly classified, resulting in a total accuracy of 91.7%. It is noteworthy the card-sort measure is a form of self-report and the validity of self-report of those who have sexually offended has been called into question. It is possible that self-report during this study was useful due to the fact that all of these individuals admitted to their sexual interests and offending behaviors and thus were already open, a typical first goal of sex offender-specific treatment. Consequently, this study may reflect some bias as it is not truly representative of the general child offending population. Instead, PPG may be utilized with offenders who do

not admit to the offenses (but have been convicted in a court of law) because if offenders do admit to sexual interest in children, psychologists may already have the necessary information regarding subjects' sexual arousal. A final problem with this study is that researchers assume all child offenders are actually pedophiles, whereas other studies (and the offending population in general) cannot be sure subjects are pedophiles, given their non-admission and sexual interactions with individuals of various ages.

Hall, Proctor, and Nelson (1988) compared arousal of 122 males who offended against individuals of different ages (children up to 12 years, minors up to 17 years, and adults). These subjects resided in a state hospital, but were not diagnosed as psychotic and did not take psychotropic medication. Stimuli consisted of audiotapes involving sexual intercourse with a consenting adult female, intercourse with a willing female child, rape of an unwilling female child (child sexual assault), and nonsexual violence against a female child (child nonsexual assault). Tapes involving children were four minutes and the tape involving only adults were two minutes and all scenarios were presented to subjects in each of two sessions. The maximum response score for each depiction was averaged. Researchers did not find any significant differences in the responses between subject groups. Consequently, researchers combined the groups to study differences between subject responses to the particular stimuli. They found nondeviant stimuli were significantly different than responses to all other stimuli. Subjects evidenced most arousal to stimuli involving a willing child, followed by consenting adult interactions, child sexual assault, and finally child nonsexual assault. Though it is noteworthy subjects were most aroused to stimuli involving a child, it is unclear what the response of a control population would be when exposed to the same stimuli. That is, subjects' arousal to child

stimuli suggests the possibility they may experience deviant arousal but this arousal pattern may also be due to differences in the stimuli that are not necessarily associated with deviant arousal, to which “normals” may respond as well.

Blanchard, Klassen, Dickey, Kuban, and Blak (2001) also compared differences in PPG responses of 324 males in an outpatient clinic who offended against unrelated children (child molesters), related children (incest offenders), and unrelated adults (adult offenders). Additionally, all subjects indicated they were primarily sexually interested in adult females or equally interested in adolescent females 15 to 16 years of age and adult females. Stimuli consisted of slides presented during narratives (recorded second person in a male or female voice matching the subject of interest) in one of six categories of sexual interactions: prepubescent girls, pubescent girls, adult women, prepubescent boys, pubescent boys, adult men. Additionally, nonsexual neutral stimuli were presented, recorded in both male and female voices. Responses were measured as largest penile volume during the stimuli and the area under the curve during the stimulus presentation. Researchers defined a pedophilic index as the response to the highest deviant stimulus minus the highest adult sexual stimulus. Adult offenders were grouped according to the number of sexual partners they have had overall (including their victims) because it is assumed this combination captures sexual interest in adults (which was born out in this study as increased number of adult partners was associated with increased sexual arousal to adults).

As with many previous studies, results of this study suggest that PPG responses differ amongst groups of offenders and control subjects. For child molesters, sexual response increased as the number of child victims increased and when there were three or

more victims, adolescent arousal tended to be higher than adult arousal. In the incest offender group, subjects with three or more victims evidenced significantly lower arousal to adults than do subjects with one or two victims. In the child molester group, specificity in men who had three or more victims was 61% whereas sensitivity in adult offenders with 24 or more sexual partners/victims was 96%. It is noteworthy that this low rate of specificity may reflect child molesters offending against children for reasons other than sexual arousal to children, such as availability of the victim or inability to obtain consensual sexual interactions with age-appropriate individuals. Another possible reason for this low rate of specificity may be problems with the stimuli, as Chaplin, Rice, and Harris (1992) were able to differentiate between control and offending groups with 100% specificity and sensitivity with very different stimuli.

PPG as a Predictor of Sexual Recidivism

Hanson and Bussière (1998) conducted a meta-analysis of 61 studies predicting sexual, nonsexual violent, and general recidivism. Studies included a mixed group of adult offenders sampled from institutions (psychiatric and penal), the community, and both of the aforementioned settings. Studies also utilized various methods of defining recidivism including reconviction, arrests, self-reports, and parole violations. Hanson and Bussière (1998) independently rated the adequacy of recidivism information in the studies and found all to be acceptable, with none relying exclusively upon self-report and many including more than one source of information. Studies resulted in information on 28,972 offenders and all studies utilized a matched longitudinal follow-up design. Only one finding of each type of predictor variable was used from each study and researchers utilized the variable that best represented the area of interest. Researchers used many

predictors of recidivism in the following general areas: demographic, general criminality, sexual criminal history, sexual deviancy, clinical presentation and treatment history, developmental history, psychological maladjustment, and other psychological problems. However, in order to be included in this meta-analysis, each specific factor had to be included in at least three studies. Hanson and Bussière utilized a median correlation coefficient and a weighted average correlation coefficient (to adjust for varying base rates amongst studies) prior to aggregating the results. The strongest general predictor for sexual recidivism was sexual deviancy and the strongest single predictor for sexual recidivism was sexual interest in children, as measured by phallometric assessment. Sexual interest, as measured by different viewing time measures, were not included in this study, suggesting more research needs to be done in utilizing viewing time measures to predict sexual recidivism. Finally, it is important to note that, due to the inclusion of so many variables to predict sexual recidivism, some predictors will appear to be significant, though this will be an error, but this will be less likely as the number of studies for each predictor increases. For the sexual interest in children utilizing PPG, there were seven studies, consisting of 4,853 subjects, suggesting this correlation is not due to chance and could be replicated.

Hanson and Morton-Bourgon (2005) conducted a meta-analysis of 82 recidivism studies, including only 35 of the original studies and containing updated information on 10 of the studies in the previous meta-analysis (Hanson & Bussière, 1998) with 29,450 sexual offenders. The remaining 37 studies were new studies to the meta-analysis. These studies involved a mixed group of sexual offenders, most of whom had been in institutions (psychiatric or penal) at some point. The general categories examined

included sexual deviancy, antisocial orientation, sexual attitudes, intimacy deficits, adverse childhood environment, general psychological problems, and clinical presentation. The authors noted this study attempted to focus more on dynamic variables predicting recidivism, so they may be addressed in treatment and assessed subsequent to treatment completion, rather than the many static variables examined in the previous study (Hanson & Bussière, 1998). The average follow-up time for the studies was five to six years. Researchers used standardized mean difference, d , as it is less influenced by base rates than are correlation coefficients. In examining the studies, effect size was not related to the thoroughness of the tracked recidivism information and was the same whether recidivism was measured by arrests or convictions. The strongest predictors of sexual recidivism were sexual deviancy (which was comprised of sexual preoccupation and deviant sexual interests) and antisocial orientation. Researchers assessed deviant sexual interests via self-report, PPG, offense history, and structured clinical ratings. The d statistic for PPG was 0.24, which was not the highest predictor of sexual recidivism found in this study. However, it is important to note that there were more studies and more subjects in the PPG studies than the other contributors to sexual deviancy. Indeed, the studies of PPG included approximately twice as many subjects as the next largest set of studies. Consequently, the d statistic for PPG is likely closer to the actual population statistic than the results for other variables. Nonetheless, it remains important to note PPG was a significant predictor of sexual recidivism, suggesting it is an essential tool in assessing risk for sexual recidivism. As noted in the discussion of the Hanson and Bussière (1998) study, measures of viewing time were not included in this study.

Consequently, this dissertation will evaluate viewing time as a potentially useful alternative, or supplement, to PPG.

Criticisms of Penile Plethysmography

O'Donohue and Letourneau (1992) discussed many problems with penile plethysmography. They noted that there “does not appear to be a standardized penile tumescence assessment, but rather there is a family of procedures which share certain common aims and features.” The authors noted the following problems with procedural variations: type of gauge used (circumferential or volumetric), type of stimuli (audiotapes, slides, or videos), content of the stimuli, duration of the stimulus presentation, length of interstimulus intervals (a fixed amount of time versus full detumescence), nature of stimulus categories (age versus Tanner stages), number of categories sampled and of the stimuli used for each category, instructions to subjects, whether or not warm-up is used, number of assessment sessions, type of recording instrumentation, whether calibration is used, data sampling rate (continuous versus multiple time intervals), whether methods are used to assess for dissimulation, gender and other technician characteristics, type of data transformation, lab characteristics (such as degree of privacy), type of sample, and setting. These criticisms raise concerns regarding the reliability and validity of the penile plethysmograph. Nonetheless, it is noteworthy that many studies utilizing PPG to evaluate sexual arousal differences between offenders and controls, and predicting sexual recidivism have yielded similar results. Consequently, the aforementioned concerns suggest the utility of PPG may be significantly enhanced with standardization of procedures and interpretation of PPG.

Howes (1995) conducted a survey of agencies and practitioners who utilize PPG in their practices. This researcher mailed 168 questionnaires to PPG assessment centers in the United States and Canada and obtained a response rate of 31%. Researchers noted a high degree of inconsistency in testing methods, stimuli, instructions, and data interpretation. As an example, methods of assessing attention to stimuli were varied and agencies used differing amounts of methods to assess for attention. Interestingly, 89% of labs used different size gauges based upon pre-test measurement. Though this suggests that many labs utilize similar methods, it is concerning that 11% of the laboratories do not alter the size of the gauge based upon subject penile circumference, which could confound test results due to significantly increased discomfort during testing. Though all agencies measured arousal to homosexual and heterosexual pedophilia, 20% did not include adult rape stimuli, 15% did not include consenting adult heterosexual activity, and 28% did not include consenting adult homosexual activity. This exclusion is critical, as the overall arousal pattern appears to significantly improve classification of group membership and prediction of sexual recidivism, which may not be thoroughly assessed without consenting adult sexual activity. A further concern related to validity is that only 42% of agencies used the same number of slides per theme and there was variability even within the same agency of exposure time to particular themes. Additionally, many agencies created their own stimuli and those who did not obtained stimuli from between five and seven different sources. Agencies were also quite variable with data interpretation such as determining what is an erection and different methods of converting data for analysis (e.g., raw scores, percent of erection, whether subjects are requested to produce a full erection, how full erection is defined). Indeed, results of this

study are highly consistent with the O'Donohue and Letourneau (1992) literature review, suggesting there is not only significant variability in PPG in research, but also in practice throughout North America.

Murphy, Haynes, Coleman, and Flanagan (1985) studied the ability of nonoffending males to alter their sexual arousal as measured by the PPG or “fake” the results of PPG. The study consisted of 203 males recruited from the community. Stimuli were two minutes in length depicting rape and heterosexual stimuli; films of lesbian interaction were also included to control for response due to novelty. There were two lesbian scenarios, two heterosexual scenarios, and four stimuli depicting rape. Subjects were presented stimuli under two conditions: one in which they were told to become aroused to a particular stimulus and one in which they were told to not become aroused to a particular stimulus. Researchers converted scores to percentage of full erection (which was determined by subject self-report either during stimuli or a heterosexual film at the conclusion of the study). When provided instructions to become aroused to “mutual” (heterosexual and lesbian) stimuli, subjects evidenced their highest achieved arousal (58%). Subjects were also able to suppress some arousal to the aforementioned stimuli when instructed to do so and achieved 44% of full erection. When instructed to produce arousal to rape stimuli, subjects only produced 37% of full erection, whereas they produced 40% of full erection when attempting to suppress arousal to rape stimuli. Consequently, there does not appear to be a significant difference in subjects' responses to rape stimuli, regardless of instructions. This lack of change may reflect an inherent inhibitory response to rape scenarios in nonoffending males, resulting in subjects being unable to further suppress response. Additionally, these results also raise the possibility

that nonoffending males are unable to enhance their responses. If this is indeed the case, the arousal to mutual stimuli under the arouse condition may reflect true arousal, rather than an attempt to enhance arousal. Additionally, it would raise the possibility that the difference between the “arouse” and “suppress” conditions to “mutual” stimuli reflects an ability to suppress, rather than enhance, arousal. The aforementioned hypothesis is consistent with a non-discriminating arousal pattern often seen in males who sexually offend against children. Overall, however, results of this study raise the possibility of suppressing arousal on the PPG, which can change an offender’s overall sexual arousal pattern, complicating prediction of future sexual offending behavior and raising concerns regarding the validity of PPG.

McAnulty and Adams (1991) studied the ability of nonoffending males (both with and without incentives) to control their sexual arousal in a variety of conditions to explore their ability to dissimulate arousal. Both groups consisted of 16 subjects and the study was later replicated with a group of homosexual males. Three color films were presented that all included foreplay, oral-genital contact, and sexual intercourse and varied with regard to the participants depicted (heterosexual, homosexual male, and homosexual female). Three participants were excluded from the study as they did not evidence sexual arousal to any of the aforementioned videotapes. Experimental stimuli consisted of eight audiotapes of heterosexual and homosexual sexual activity (four of each) and subjects were exposed to the stimuli with both standard and “faking” instructions. Color slides of nude individuals who were similar to the individual depicted in the audiotape were shown concurrently as well. Another variable of interest was subjects’ response to a “signal detection” task during stimulus presentation (which was

included in some of the stimuli). Additionally, all subjects were presented with a recognition task that involved a multiple choice test with questions ranging from the color of a depicted person's hair to the partner's age as described in the narration. Specifically, there were four conditions varied by standard versus faking instructions and with or without a signal detection task (in this study, subjects were to press a button when they heard terms referring to male genitalia). Unlike some other studies, the faking instructions only involved suppression of arousal, rather than also including enhancing arousal to any stimuli.

Some results of the McAnulty and Adams (1991) study were somewhat unexpected. Specifically, the signal detection task resulted in less arousal overall regardless of the instructions. These results raise the possibility that engagement in the signal detection process actually distracts from stimuli, resulting in less penile responsiveness. Subjects also responded significantly more to stimuli during standard instruction segments than faking instructions, which is consistent with the results of other studies indicating subjects are able to voluntarily inhibit their sexual arousal; this finding was also replicated with homosexual subjects. Subjects who were most able to suppress responses also tended to evidence the lowest scores on the recognition task, suggesting that suppression did slightly interfere with recognition. Although subjects on average were able to significantly suppress their responses, suppression was not strong enough to alter their overall sexual arousal pattern. That is, the overall arousal was decreased but the shape of the arousal pattern remained the same. Additionally, several subjects were able to completely suppress their arousal. These subjects are especially interesting, as they may provide insight into further ways to detect dissimulation or even prevent suppression

of sexual arousal. Unfortunately, discussion of the subjects' strategies to suppress arousal was not discussed in this study. Finally, though the signal detection task did not reduce subjects' ability to suppress arousal, it may be that the task was too simple or there were not enough signals. An offender who is sophisticated and interested in altering his arousal pattern may attempt to suppress arousal to children while increasing his arousal to adults, and this study at least bears out the possibility of arousal suppression. Indeed, this study may help to explain the non-discriminating sexual arousal pattern seen in studies of offenders.

Adams, Motsinger, McAnulty, Moore (1992) also compared sexual arousal and ability to alter responsiveness during PPG assessment. Researchers studied nonoffending heterosexual and homosexual males, as much research has supported an ability to discriminate between those arousal patterns. Subjects were shown three videos, depicting heterosexual, homosexual male, and homosexual female sexual interactions including fondling, oral-genital contact, and sexual intercourse. If subjects showed a response to heterosexual and homosexual female stimuli, they were included in the heterosexual group; males included in the homosexual group were required to show arousal to the homosexual male stimuli and decreased arousal to homosexual female stimuli. Subjects who did not arousal to any stimuli were excluded from the study.

Stimuli consisted of four total audiotapes consisting of either heterosexual or homosexual males (two of each) concurrent with slides of male or female nudes (appropriate to the audiotape played). After each stimulus subjects were asked to rate the degree of their sexual arousal (cognitive) and tumescence on a scale from zero to 10. Subjects were also asked to describe the previous scenario (both audio and visual) in

detail in order to determine their attention to the stimuli. During the first two stimuli presentations, subjects were provided standard instructions; they were provided faking instructions in which they enhanced their arousal to non-preferred stimuli and suppressed their arousal to preferred stimuli prior to the last two stimuli. Both groups were significantly able to suppress tumescence to the preferred stimuli, but could not enhance their responses to the non-preferred stimuli. Additionally, subjects were accurate in their appraisal of their ability to both suppress and enhance their arousal. Interestingly, both groups also believed they were more able to control their tumescence than their cognitive arousal. This perception raises the possibility a cognitive measure of sexual interest, such as viewing time, may be useful as a replacement or supplement to PPG. It is also noteworthy that there were no significant differences in recall between the standard and faking conditions, suggesting that faking may not be detected via this recall method. However, it remains possible that the attention to detail that was required of subjects increased the subjects' difficulty in enhancing arousal or further suppressing their tumescence. Additionally, this inability to enhance arousal in combination with an ability to suppress arousal may explain the non-discriminating pattern found in child molesting males. Finally, it is noteworthy that though subjects were able to generally suppress their arousal significantly (though not completely), their overall responses (in context with each other) remained the same, as heterosexuals evidenced most arousal to the heterosexual stimuli and homosexuals evidenced most arousal to the homosexual stimuli. This finding further supports the need to examine general arousal patterns, rather than simply observing particular data points independently.

Quinsey and Chaplin (1988) also studied subjects' ability to alter their sexual responses. In this small study of 16 subjects, 15 from the community and one from a psychiatric institution, subjects were exposed to 18 different stimuli under four conditions. Stimuli were all narrated in the first person describing some interaction with an adult female including three nonsexual (neutral), five consenting heterosexual intercourse, five rape, and five nonsexual violent. The instructions included standard and faking, as well as with and without a signal detection task. In this study, the signal detection task involved semantic tracking, whereby subjects were asked to press one particular button when sexual activity occurred, another button when violent activity occurred, and both buttons when violent sexual activity was presented. Therefore, this task required subjects to closely pay attention to the stimuli in order to perform well on the semantic tracking task. In the faking condition, subjects were instructed to appear aroused to rape and nonsexual violence but not to consenting sex. They were also offered suggestions regarding ways to dissimulate responses but were also able to create their own. Subjects were tested over four sessions, with a different set of instructions each session.

Results of the Quinsey and Chaplin (1988) study supported the notion that the signal detection task does interfere with faking. During standard instructions with no button presses, subjects showed most arousal to consenting sex, followed by rape, and finally neutral stimuli and nonsexual violence were similarly unarousing to subjects. In the faking instructions with no button presses, subjects showed more arousal to rape and nonsexual violence than to the consenting stimuli, suggesting they were able to alter their arousal pattern. However, during faking instructions with button presses, the overall

pattern of responding returned to that found during the standard instructions, suggesting that the processing task interfered with subjects' ability to dissimulate their arousal. As researchers utilized z scores, it is unclear how subjects' absolute level of arousal may have been altered. Nonetheless, the interference with dissimulation during the button presses suggests the importance of utilizing a method of processing stimuli to increase the likelihood of detection of faking to obtain an accurate arousal pattern. Though on average subjects were unable to dissimulate their arousal pattern during the faking instructions with button presses, two subjects were significantly able to produce a "normal" profile. These subjects reported they accomplished this task by memorizing the stimuli (faking with button presses was their fourth session) resulting in an ability to appropriately time their button presses without attending to or processing this stimuli. Consequently, asking subjects attend to the stimuli resulted in an inability to fake unless they are quite familiar with the stimuli; this highlights the importance of presenting clients with stimuli that are novel to them. Finally, results of this study also highlight the importance of examining arousal levels in context of each other, rather than simply comparing individual data points independently.

Monarch 21 addresses validity concerns

One Penile Plethysmograph system, Monarch 21 (by Behavioral Technology, Inc.), has incorporated many procedural changes to address validity concerns that were previously discussed. First, Monarch 21 PPG is a standardized system, rather than a data recording device or set of stimuli. All users of the Monarch 21 have similar equipment produced by Behavioral Technology, Inc. Additionally, administration procedures are scripted and standardized; technicians all receive Monarch 21 Level One Technician

training in which they are also trained in appropriate utilization of the instrumentation, testing administration, and following standardized procedures.

Monarch 21 stimuli are also standardized. Stimuli are all presented for the same amount of time and the core stimuli all have the same format. Core stimuli are varied with regard to stimulus age (preschool, grammar school, pubescent, and adult), gender, and level of manipulation (persuasion versus coercion). All users of the Monarch 21 PPG system utilize these core stimuli as part of every evaluation. Data are continuously recorded and graphs are printed at measure completion including a variety of variables such as the maximum response during stimulus presentation and the area under the curve during stimulus presentation.

To address concerns related to attention and decrease likelihood of dissimulation, Monarch 21 includes a signal detection task. Specifically, during the visual portion of the stimuli, subjects are randomly presented with a star in a corner of the screen (which randomly varies). Measures of both attention and stimuli processing are included in the audio portion of the stimuli. Tones are included in the audio stimuli to measure for attention, as subjects are requested to press a button upon hearing the tone. Additionally, subjects are requested to press the button one time if the stimulus is persuasive and twice if it is coercive. Results of these signal detection tasks may be analyzed by the evaluator to determine if the subject was maintaining attention and processing stimuli to determine the validity of the measure. Additionally, signal detection tasks have been shown to significantly interfere with dissimulation, resulting in subjects showing a similar sexual arousal pattern during signal detection tasks as when they respond naturally to the stimuli (Quinsey & Chaplin, 1988).

Though the Monarch 21 PPG system has made vast improvements to address concerns related to standardization of PPG, there remain concerns regarding utilization of PPG. First, administration of the PPG requires a significant time investment: approximately three hours, two of which the subject is sitting in the exam room. Another drawback to PPG is the major financial investment required to purchase a system. Training to administer the PPG also requires many hours of technician time. Finally, a significant concern regarding PPG is the intrusive nature of the measure.

A measure of viewing time may address the aforementioned concerns regarding the PPG in a variety of ways. First, viewing time measures require little time to administer, typically less than an hour. Additionally, little training in administration is required of viewing time. Indeed, one particular measure of viewing time, Affinity (Glasgow & Croxen, 2003), provides testing proctors with a script to follow during administration. Measures of viewing time also do not require a significant up-front investment, allowing more evaluators to utilize this measure. Finally, measures of viewing time are not intrusive, as the subject is seated in a private room fully clothed and is simply asked to rate sexual attractiveness of individual images.

Discussion of Viewing Time

The amount of time an individual spends viewing stimuli has been studied since 1942 (Rosenzweig, 1942). It was hypothesized that length of time viewing stimuli is positively associated with sexual interest in those particular stimuli. If this is, indeed, the case and sexual interest is found to be associated with sexual recidivism, viewing time measures may either replace or supplement PPG in the assessment of those who sexually offend.

Early Studies of Viewing Time

Rosenzweig (1942) researched viewing times of 20 adult male schizophrenic patients who were hospitalized. Prior to engagement in the studied task, staff rated potential subjects on spontaneous display of sexual behavior (such as masturbation). Ratings were reviewed and subjects who fell into a “high” and “low” spontaneous sexual behavior group were included, with 10 subjects in each group. Stimuli consisted of three sets of 24 cards. Four cards were neutral practice cards, five had line designs, five had images of a neutral person or object (such as a landscape), five showed a romantic scene, and five depicted an overtly sexual scene. These cards were presented in a “photoscope” which displayed the cards in a similar fashion to a Rolodex, in which subjects could view the images as quickly or slowly as they wished, but only in the same, predetermined order. Subjects were left alone in the room to view the images as they wished, and the experimenter moved to a room with a one-way mirror in which he could view the number of the card presented. The experimenter then measured the amount of time subjects viewed each card. Subjects were allowed to go through the card set at least two times, but no more than three times. This occurred during three sessions, though only two sessions were included in this study as the second session was included as part of a larger study. That is, the first and third sessions were presumed not to be influenced by events that occurred between those two sessions, whereas the subjects were taking medication during the second session.

Rosenzweig (1942) summed the time spent viewing sexual versus nonsexual images (not including the practice cards). During the first session, both high and low sexual interest groups viewed nonsexual images for a total of 19 seconds. However, the

high sexual interest group viewed sexual cards for 40 seconds whereas the low sexual interest group viewed sexual images for 13 seconds. This suggests subjects in the high sexual interest group were more interested in sexual images than were subjects in the low sexual interest group, which is consistent with staff ratings. Additionally, this also supports the notion that longer viewing time is associated with sexual interest. During the third session, the low sexual interest group spent a similar amount of time viewing sexual and nonsexual images (17 and 18 seconds, respectively). The high sexual interest group spent 12 seconds viewing nonsexual cards and 25 seconds on sexual cards. Thus, though those in the low sexual interest group spent similar amounts of time viewing sexual and nonsexual content, high sexual interest subjects spent twice as long viewing sexual stimuli (as compared to nonsexual stimuli). Results of this study suggest that longer viewing times are associated with increased sexual interest in the images being viewed.

Zamansky (1956) compared viewing times of 20 heterosexual and 20 homosexual males. The researcher presented two images next to each other, of various subjects. Subjects were asked to judge which image was larger (though they were equal in size) while the experimenter was able to view the eye movements of the subject and timed the overall time subjects looked at each picture. Subjects were encouraged to view the images as long as necessary to determine which image was larger. Stimuli were 24 pairs of images as follows: six pairs of one male and one female, three pairs of more than one male and more than one female in each picture, four pairs in which females were “typical” but the male hinted at homosexuality, four pairs of a male and a neutral stimulus (such as a landscape), four pairs of a female and a neutral stimulus, and three pairs of neutral stimuli so subjects would not become aware of the sexual nature of the

study. For the nine male-female pairs, viewing time was scored as the number of seconds spent looking at the male pictures less the number of seconds viewing the female pictures. These scores were averaged across all stimuli for each individual and were then averaged across each group. As expected, homosexual subjects viewed male images longer (3.55) and heterosexual subjects viewed female images longer (-2.37), which was statistically significant at $p < 0.001$. Median scores were also significantly different at the same significance level. Results of this study also support the hypothesis that individuals spend more time viewing sexually interesting objects than images that are not sexually interesting to them.

MacCulloch and Sambrooks (1974) performed a single case study in which they hypothesized that time viewing a particular stimulus would change as did the subject's sexual interest. The researchers engaged in aversion therapy with a subject who was highly motivated to decrease his sexual interest in males. Researchers measured sexual interest by subject self-report, the "sexual orientation method" ["SOM"] which is a paper and pencil method, and time spent viewing an adult male image. The subject reported decreased sexual interest in males, which also was consistent with the results of the SOM and decreased viewing time. Researchers concluded that this supports the notion decreased viewing time is associated with decreased sexual interest. However, it is noteworthy that there were multiple problems with this study. First, the time viewing the images were under the subject's control and if the subject viewed the image for a certain amount of time during aversion therapy, the subject experienced an aversive stimulus. Consequently, it is possible that the viewing time was artificially decreased by fear of punishment, rather than lack of sexual interest. Additionally, a single subject design is not

robust in determining the underlying construct being measured, as it is unclear if this is generalizable to a larger group of people.

Love, Sloan, and Schmidt (1976) examined differences in viewing time of varying levels of sexual images based upon subjects' level of sexual guilt. Subjects were 41 male college undergraduate students. Subjects were first administered a measure to determine their "sex guilt" and were divided into three similar-sized groups: low sex guilt, moderate sex guilt, and high sex guilt. Stimuli were 18 black and white slides of varying levels of pornography, ranging from females in bikinis to graphic heterosexual activity. Subjects were asked to rate each image in terms of obscenity, attractiveness, disgust, and artistic value and were instructed to take the time they need to accurately rate the images, as they could not view the images again. The experimenter left the room and timed the period the subjects viewed each image. This was accomplished via a hidden sound system in which the experimenter could hear when the subject changed the slide.

Results of the Love, Sloan, and Schmidt (1976) study revealed an overall effect of the level of pornography. Researchers ranked each image based upon the average subject obscenity rating, from lowest to highest. Results showed that when considering subjects overall, viewing time increased as the obscenity rating increased. This suggests that as sexual interest increases, so does viewing time, which is highly consistent with the results of other studies.

Modern Studies of Viewing Time

Abel, Lawry, Karlstrom, Osborn, and Gillespie (1994) studied the ability of a viewing time measure (the Abel Screen) to discriminate between control subjects and subjects who have sexually offended. Researchers compared the sexual interest patterns

of admitted pedophiles with those of control subjects who are not known to sexually offend. Subjects included a control group of 101 adult males and 185 adult males who offended against age-inappropriate individuals including pubescent males and females and prepubescent males and females. Stimuli consisted of two sets of slides with images of children, adolescents, and adults. In the initial set of slides, individuals were partially clothed whereas all individuals in the second set of slides were nude. Subjects were asked to rate sexual arousal when viewing each of the first set of slides. During the second set of slides, physiological response to the slides was recorded with the subjects' awareness.

Researchers (Abel et. al, 1994) performed four discriminate function analyses, one for each group of offenders. Sensitivity was 76% and specificity was 98% when discriminating between controls and males who offended against prepubescent boys. For adults who offended against prepubescent girls, sensitivity was 91% and specificity was 77%. Discriminating between the control group and pubescent males was most effective, as sensitivity was 90% and specificity was 98%. Finally, in males who offended against pubescent females, sensitivity was found to be 86% and specificity was 77%. Review of the previous results suggests that specificity, especially with female victims, is quite low. This is especially concerning as the researchers recommend this measure to be used for initial screening. When testing the equation developed via the previous discriminate function analysis, specificity of offenses against female victims fell another 8.9%. Given the initial specificity, this result is quite poor. Additionally, the researchers did not provide any information regarding the sexual interest pattern or variables or the equation used to discriminate between groups. Consequently, these results may not be replicated. Finally, it appears that the nude images contributed to better discrimination between

groups. However, researchers no longer use nude images, due to ethical and legal concerns. Though the specificity of this screening measure was poor, there does appear to be at least some association between length of viewing time and sexual interest. However, this study demonstrates a critical concern with the Abel group of viewing time measures. Specifically, research is conducted predominantly in Dr. Abel's lab, raw data is not provided to other researchers, and little necessary information for replication (such as the discriminate function equations) are provided.

Wright and Adams (1994) compared viewing time of 80 college students who varied by sexual orientation and gender. There were four groups of 20: heterosexual males, homosexual males, heterosexual females, and homosexual females. Stimuli consisted of 60 slides including 20 nude males, 20 nude females, and 20 neutral stimuli. Each slide had a white dot in one corner or the center of the image, which was varied across slides and groups. Subjects were asked to press a button on a box that was analogous to where the white dot appeared and to do so as quickly and accurately as possible. Measured viewing time was the interval between stimulus presentation and the response regarding the dot placement. Additionally, whether the response was correct or incorrect was monitored. After the previous task was completed, subjects were again exposed to the stimuli, in order, and asked to press the button where the white dot had been (an incidental learning task). Data with an initial incorrect response were not analyzed.

Results of the aforementioned study (Wright and Adams, 1994) were consistent with expectations. Homosexual males spent significantly longer viewing images of males than females or neutral stimuli. Heterosexual males viewed female slides significantly

longer than male slides. Homosexual females viewed female slides significantly longer than they viewed male slides, and heterosexual females viewed male slides significantly longer than they viewed female slides. These results are highly consistent with other studies of viewing time, as individuals tended to look longer at slides of people to whom they are generally attracted. When presented with male slides, homosexual males took significantly longer to respond than did heterosexual males and homosexual females; heterosexual females did not significantly differ from the other groups. When responding to female slides, heterosexual males and homosexual females took significantly longer to respond than did heterosexual females. Homosexual males did not significantly differ on female slides from any other group. Results of this study suggest that sexual orientation, and thus sexual interest, interfered with the choice reaction time task, as subjects all looked longer at preferred images. These results further support the use of viewing time as a measure of sexual interest.

Harris, Rice, Quinsey, and Chaplin (1996) studied sexual interest as assessed via viewing time and penile plethysmography. Subjects were 26 adult males who offended against children under the age of 14 and 25 control subjects who were not known to sexually offend. Of the child molesters, six also offended against adult women. Individual subjects did not complete all measures. Ten control subjects completed a rating task and the viewing time measure, and 15 control subjects completed the PPG. Eleven offenders completed the rating and viewing time task and the remaining 15 offenders completed the PPG. The stimuli for the rating and viewing time measure consisted of 70 slides with 10 slides in each category that varied by Tanner stage (1, 3, and 5) and gender as well as a neutral category. Subjects were asked to carefully review

each slide because they would be questioned about the slides later. After reviewing all the slides, the tray was reset and subjects were asked to rate the attractiveness of each slide. PPG stimuli included the same seven categories of images as the viewing time slides, but there were only two stimuli per category, one of which was clothed and the other of which was nude. As is typical with PPG, the maximum arousal change during stimulus presentation was used as a measure of sexual arousal. Three dependent measures were used for the viewing time assessment including mean category ratings, mean category viewing time, and a deviance index (the longest mean viewing time for adults less the longest mean viewing time for children or adolescents).

Visual inspection of the results of the aforementioned study (Harris, et. al., 1996) showed similar responses for ratings, viewing time, and PPG with the control group who were not known to sexually offend. Inspection of PPG, viewing time, and ratings in the offender group appeared to be different. Significant differences were found between the control and offender groups, suggesting that the overall profiles were different. Additionally, the control group had longer overall viewing time than did the offender group. It is possible that the longer viewing time is because the control group was open with their responding whereas the offender group members were attempting to present a positive image of their sexual interests or were trying to dissimulate the results. Significant differences between groups were found in the deviance differentials of both the PPG and viewing time measures. Overall, results of this study showed there were significant between group differences on both the viewing time and PPG measures. However, based upon statistical analyses, PPG discriminated between groups better than

did the viewing time measure (common language effect size of 0.93 versus 0.80, respectively).

Quinsey, Ketsetzis, Earls, and Karamanoukian (1996) studied the differences in viewing times in two studies, one with adult males and females and one with only adult males who took a PPG. The first study had 48 subjects (24 male, 24 female). Subjects were provided with 36 experimental slides (18 male, 18 female), one landscape slide, and one final slide indicating the slides are completed. Images included males and females who were children, pubescents, and adults all nude and with a full frontal view. Subjects were asked to study the slides so they could answer questions about them later; they were then asked to go through the slides again and to rate the attractiveness of the people in the slides. Subjects rated members of the opposite sex and adult age significantly higher than all other ratings. Additionally, subjects viewed images of the opposite sex longer than individuals of the same sex and the length of viewing time decreased with the age of the stimulus. For males, the correlation between ratings and viewing time was 0.80 and the correlation between ratings and viewing times for females was 0.60. Though these correlations are quite high, it is noteworthy that subjects in this study had no reason to dissimulate their sexual interests. Additionally, research has shown that the relationship between control subjects' ratings and viewing times is more consistent than the same relationship found in a group of offenders (Harris et. al, 1996).

In their second study, Quinsey et. al (1996) compared results of a viewing time measure and PPG. Subjects were 24 heterosexual males recruited from the community. Stimuli were 18 slides, two of which were neutral. The remaining 16 slides, two from each category, were of nude males and females of various age categories (children,

prepubescents, pubescents, and adults). Additionally, subjects were shown a five minute video of an adult male and female engaging in consensual sexual intercourse as a “warmup” exercise followed by a 60 second exposure to each of the 18 slides, with a minimum of a 30 second interval between stimuli to allow return to baseline. During the viewing time task, subjects were asked to rate the sexual attractiveness of the individual to them. Viewing time was measured as long as each slide was illuminated. Researchers averaged the responses for each category for both PPG and viewing time. Subjects evidenced significantly more arousal to adult females than to any male category, neutral stimuli, and to the female child and female prepubescent categories. Though there was a trending difference between arousal to female adults and female adolescents, response to the female adults was not significantly greater than arousal to female adolescents. However, arousal to female adolescents was significantly higher than the response to any male category. Subjects rated adult females significantly higher than all other categories and no difference was found between ratings of male and neutral stimuli. Additionally, subjects spent significantly more time viewing female than male slides; subjects also viewed adult female slides longer than any other female slides. Researchers performed correlational analyses between viewing time, ratings, and PPG for each subject. Of the 24 correlations between viewing time and PPG, 12 were statistically significant and the range was from -0.05 to 0.84. Results of this second study further supported the notion that length of viewing time is positively associated with sexual interest. Finally, though PPG and viewing time were directly compared, which is highly unusual, neither measure was a standardized method, but were only developed for this particular study and are not

commercially available. Consequently, these data do not necessarily represent measures that are typically used in clinical practice.

Abel, Huffman, Warberg, and Holland (1998) examined sexual arousal and interest adult males who admitted to offending against children. Subjects were 157 adult males who referred to an outpatient clinic after been accused of sexual offending behaviors; subjects who did not admit to sexual offenses were excluded from the study. Though this exclusion produces a more homogenous group for study, it is a flaw as these deniers are of most interest for sexual interest and arousal testing. All 157 subjects were administered the Abel viewing time measure, though only 56 subjects completed the PPG. This is yet another study flaw, as the larger number of participants increases the power of the Abel test as compared to the PPG. PPG stimuli consisted of landscapes, and males and females of the following ages: four years, eight years, 12 years, 16 years, and 22 years. The Tanner stages of these individuals were not discussed; consequently, the Tanner stage and developmental level of the 12-year-olds remains unclear. That is, a 12 year old may look like an undeveloped child, a prepubescent, or a pubescent. Additionally, all individuals depicted were nude. Subjects' baseline was measured for three minutes prior to initial stimulus presentation. Slides were shown for one minute and the minimum interval between stimuli was 30 seconds and was extended until the subject returned to baseline of five mm within baseline. There were 40 total slides, though it is unclear how many stimuli were in each category. Stimuli for the viewing time consisted of seven slides in each of six categories (male and females from 8-10 years, 14-17 years, and over 22 years). All models wore bathing suits in the pictures. Other than these six categories, other stimuli varied, but were not included in the analysis. Researcher

reviewed the viewing time data and “disparate responses... were removed prior to data analysis.” No explanation for how or why this was done was provided in the study. This is another criticism often directed at the Abel viewing time measures.

Review of the results showed that PPG reliability was generally similar to, or better than, viewing time reliability (Abel et. al, 1998). The only instance of PPG reliability being less than viewing time reliability was in the “young male” category; for PPG, this age range included children two to four and eight to 10 years of age. All other PPG reliability coefficients ranged from 0.85 to 0.97; viewing time reliability coefficients ranged from 0.84 to 0.90. Subjects were divided into five groups, based upon victim characteristics: “other” (offenses against adults including rape, exhibition, and voyeurism), males less than 14 years, females less than 14 years, males between 14 and 17 years, and females between 14 and 17 years. Researchers used the “other” group as control subjects. As previously discussed, using an offender group as a control group is problematic as it is not clear how these adult offenders may differ from non-offending subjects. Additionally, treating these subjects as controls suggests a degree of “normality” which is problematic when considering any person who has sexually offended. Researchers performed a correlational analysis between the victim characteristics (age and gender) and the stimuli associated with that group. The viewing time correlations were significant for male and female children as well as adolescent males. When researchers excluded PPG non-responders, no significant correlations were found between responses and victim choice; when all subjects who took the PPG were included, the correlation was significant for male children and adolescent females. It is noteworthy that researchers did not report the particular correlation coefficients, but

simply reported whether or not it was significant. Finally, when comparing the each measure's ability to predict group membership (based upon victim characteristics), no significant difference was found between viewing time and PPG for predicting offenses against female and male adolescents as well as male children.

Fischer and Smith (1999) had several criticisms of the above study. First, they noted that it was not clearly articulated in the study if the subjects had varied victims (in terms of age and gender) or if they had more specific targets. They also stated that Abel et. al (1998) did not describe how the testing scores identified particular target victims. Fischer and Smith criticized Abel et. al for not reporting the Pearson product-moment correlations and having only two adolescent males in the norms, though Abel et. al has indicated this measure may be used with adolescents. Authors also noted that Abel et. al did not explain the rule by which target victims were classified into categories using a subset of slides, there was a significant correlation with a past victim (though the correlation was not specifically reported). Finally, Fischer and Smith noted that a series of studies on this measure did not provide validity for the 22 stimuli categories, as the only stimuli overlap between studies were in six categories and the hit rates were not reported.

Abel, Jordan, Hand, Holland, and Phipps (2001) examined viewing time patterns of a variety of offending individuals. Subjects were 747 adult males who had been administered the Abel Assessment for sexual interest between 1994 and 2000. Subjects were first split into two groups: one to build the model of sexual interest pattern for different groups, and another to test the model. These participants were further divided into three groups, based upon their sexual behaviors: non-child molesters (denied interest

in children, not accused of offending against children, and denied molesting children), admitting child molesters (admitted to offending against at least one child under 14 who was not a family member), liar-deniers (who had been accused by more than one family, convicted of child molestation, or therapist believed the client's excuse for why he was wrongly accused was unreasonable). Stimuli consisted of 16 types of images including both genders, Caucasian and African American, and individuals aged two to four, eight to 10, 14 to 17, and over 21 years of age. Models all wore bathing suits and were framed against a plain blue background. Researchers utilized stepwise regression analyses for each group of child offenders (girls under 14, boys under 14, and liar-deniers). These models were then utilized with the "holdout" subjects. Among other variables, males who offended against females under 14 years spent more time viewing eight- to 10-year-old females and rated eight- to 10-year-old females higher than the "control" group did. When the model was tested on the holdout group, one cut point had a sensitivity of 74% and specificity of 73%; to obtain a specificity of 99%, sensitivity was reduced to 25%. As expected, increased time viewing eight- to 10-year-old males was associated with having offended against males under 14 years of age, amongst other variables. When the model was applied to the holdout sample, one cut point produced a sensitivity and specificity of 86%. With a specificity of 99%, sensitivity was reduced to 28%. Finally, researchers examined variable patterns associated with liar-deniers, who are arguably of most interest when attempting to discern sexual interest in offenders. The regression model included specific hobbies and interests, higher of the two former predicted values from the aforementioned models, and the "Behavior Denier Scale" which is comprised of inappropriate sexual behaviors. When the model was applied to the holdout sample, one

cut point had a sensitivity of 75% and a specificity of 76%. In order to obtain a specificity of 96%, sensitivity was reduced to 33%. Review of the models' ability to discriminate between offender groups was significant. However, the specificity was generally quite poor, and if the specificity was increased, the sensitivity of the measure dropped to virtually useless levels. Consequently, though it appears progress has been made in developing the Abel measures of viewing time, more work needs to be done in order to be widely used and accepted. Additionally, many critiques of this measure are relevant in this study as well, including lack of adequate information needed to appropriately replicate results.

Zabarauckas (2001) studied the ability of a viewing time measure to correctly categorize offenders by their offense behavior. Subjects were 26 adult males, 19 of whom were classified as pedophiles and seven of whom raped adults. All subjects were convicted of sexual offenses and none had been placed in a federal prison. The researcher compared the viewing time mean of each category with the 15 remaining categories; if the mean of a particular category was one standard deviation or more higher than the mean of the other categories, the researcher considered the single category to reflect sexual interest. Utilizing the aforementioned method, the research correctly classified 85% of the pedophiles and 50% of the rapists. As the researcher utilized a viewing time measure from the Abel family of viewing time measures, and she did not report otherwise, it is highly likely that "trimmed" data were used. As previously discussed, use of "trimmed" data, without explanation, does not appear to be an acceptable method of treating data for research purposes.

Criticisms of the Abel Family of Viewing Time Measures

Smith and Fischer (1999) examined the test-retest reliability and discriminative validity of the Abel Assessment for Interest in Paraphilias (AAIP) with adolescent males. Subjects included 81 adolescent males in residential or day treatment programs, 41 of whom had sexually offended. Males who sexually offended did not appear to specifically target victims of any specific age or gender, as there was considerable overlap between victim characteristics. Specifically, 12 subjects offended against females of both age groups (zero to six and seven to twelve), five offended against males of both ages, nine offended against both genders in the youngest age group, 12 offended against both genders in the older age group, and four offended against all four groups. Researchers administered the AAIP to participants on two separate occasions, with an average of 14.8 days apart and computed a correlation coefficient for each subject. The correlation coefficients were then converted to Fisher's z score, all Fisher's z scores were averaged, and the Fisher's z score was reverse transformed to a correlation coefficient. The average correlation coefficient was 0.63, with a range of -0.09 to 0.96. Researchers also compared the classification of deviance across both test administrations, using both Abel's highest z score rule (sexually interested in single category with the highest z score) and his rule of thirds. For the rule of thirds, clinicians are to separate the space between the lowest child and the highest adult/adolescent scores; if a child's score crosses the lower third of this identified space, the subject is considered "deviant." Using the highest z score rule, the classification of deviance between the first and second test administrations was 0.58. The rule of thirds method was less, at 0.48. These scores suggest that the test-retest reliability for the AAIP is quite poor over a two week time interval, even in classification.

Researchers also used a chi-square analysis to examine the measure's ability to discriminate between groups (sexual offenders and control subjects). The chi-square coefficient was not significant for either the highest z score or the rule of thirds. Overall efficiency of prediction using the highest z score rule for the first test administration was 52%, and most subjects were classified as nonoffenders; the efficiency of prediction using the highest z score rule for the retest was 56%. Though the specificity was high, the sensitivity was poor. Using the rule of thirds, the overall efficiency was 59% during the first test and 58% at retest. Though the efficiency of both methods is similar, sensitivity was acceptable, but the specificity was poor for the rule of thirds, in contrast to the opposite problem using the highest z score rule. Finally, when the researchers compared the highest z score to victim characteristics in the offending sample, there was only one subject in which the offense and measure were concordant. Using the rule of thirds when examining the correlation between the deviant prediction and victim characteristics, no significance was found once the researchers corrected for error. Overall, results of this study suggest that the AAIP does not possess adequate test-retest reliability and does not perform significantly better than chance in predicting victim characteristics.

Fischer and Smith (1999) explored issues associated specifically with the Abel Assessment for Interest in Paraphilias (AAIP) and the Abel family of measures in general, which have many of the same problems. Several of the criticisms were related to the treatment of the data. Researchers noted that outliers are removed ("trimmed") prior to data analysis, but no formula for how this is done is provided to clinicians or researchers. This possibly clouds the underlying nature of the data, as those outliers may be relevant. Removal of those "outliers" also changes the measure for each individual, as

the particular slides that are removed vary each time the test is administered. The Abel family of measures utilizes ipsative z scores in which each subject is only compared to himself. Clinicians are given z scores, but not means or standard deviations for the measure. As the z scores are ipsative, rather than norm-referenced, it is unclear if the scores are distorted due to lack of variability or excessive variability. One subject may not be compared to another and, though one elevation may appear quite high, the difference may be quite small, which may not be determined without knowing the underlying data. Additionally, as the z scores are ipsative and the underlying means and standard deviations are not provided, the categories may only be considered ordinal. Nonetheless, the Abel family of measures uses language suggesting that the z scores are norm-referenced, sometimes referring to differences as “slightly higher” and “exceedingly high.” Additionally a z score of 0.01 has been called “high” and the phrase “not exceptionally high” included z scores of 0.15, 0.30, and -0.02. these descriptors are inconsistent with the z scores to which they refer.

Consistent with the aforementioned concerns, Fischer and Smith (1999) noted that the bar chart presented with the AAIP is misleading. Specifically, this measure presents a bar chart with $-3z$ on the far left and $+3z$ on the far right. All categories begin at the far left, $-3z$, and continue to their corresponding z score. The appearance suggests more similarity than would be expected had the chart centered at a z of zero and values were represented in different directions from the zero. Authors reported the rule of thirds is expressed as if it is norm-referenced; however, they noted this rule is arbitrary, does not appear to have any research supporting it, and is not actually norm-referenced. They indicated these issues may be corrected with weighting the ranks so they may be

compared, despite their ipsative nature, but this is not done in the Abel family of viewing time measures and the data is not available for the clinicians to correct. Fischer and Smith also noted that the removal of outliers while testing internal consistency inappropriately inflated the reliability. They also wrote of the problem of utilizing internal consistency as a measure of reliability, rather than also examining test-retest reliability. Researchers indicated that in a test of ability to dissimulate, test-retest reliability ranged from 0.56 to 0.74. They also noted that for this study, only 10 categories were used, rather than the full 22 categories that comprise the AAIP. Finally, Fischer and Smith indicated that in a study to support the measure's ability to differentiate between control and child molester samples, different stimuli were used than in the current instrument. Consequently, data from that study may not be applied to the current instrument, which does not have the research.

Fischer and Smith (1999) recommended a variety of ways to deal with the multiple criticisms they have of the Abel family of viewing time measures. However, the raw data is not available for study and future research outside of the developer's lab (with the exception of the Letourneau study). Consequently the suggestions may not be implemented and other researchers may not study and replicate the results of studies of the Abel family of viewing time measures.

Introduction to Affinity Viewing Time Measure

Glasgow, Osborne, and Croxen (2003) developed a measure of viewing time for learning disabled sexual offenders, called the Affinity. This measure involves ranking the attractiveness (or lack thereof) of categories of individuals (varied by age and gender), rating the sexual attractiveness of images of multiple people, and a covert measure of

viewing time while subjects make the ratings. This measure will be discussed in more detail in the Methodology section below.

Glasgow and Croxen (2003) studied some of the statistical properties of the Affinity. Subjects were 31 adult males not known to sexually offend (control group) and 31 adult males in a maximum security psychiatric facility who were convicted of offending against children 16 years or younger. Offending subjects also had to be at least five years older than their victims. The internal consistency of this measure was good, as the Cronbach alpha scores ranged from 0.76 to 0.93. Researchers ran a correlational analysis between subject ratings and length of viewing times. In the control group, 29 of the 31 correlations were statistically significant, with 28 of them significant at the $p < 0.01$ level. However in the offender group only 23 of the 31 correlations were significant; 16 of those were significant at $p < 0.01$. These results suggest that offenders are either not as open regarding their sexual interests or are not as aware of their sexual interests as are non-offenders, further supporting the need for utilizing more covert measures to detect sexual interest. Researchers transformed viewing time scores and input them into a 2X8 ANOVA (group by stimulus category). The control group showed significantly longer viewing times to adult females and adolescent females than did the offender group and this was the only significant between group difference. Both groups spent significantly more time viewing female (as opposed to male) stimuli. Researchers also computed a deviance differential where they subtracted the highest child viewing time from the highest adult viewing time and computed a ROC analysis. The area under the curve for this analysis was 0.87 and at one cutoff 96% of the offenders were correctly classified. However, there were 23% false positives. When considering the cutoff, it appears that in

control subjects have a greater discrepancy in viewing times between adults and children. That is, though both groups spend more time viewing adults than children, the control groups' time differential is larger.

Affinity addresses validity concerns

Research has shown that length of time spent viewing images is positively associated with sexual interest in similar individuals. However, there are a number of problems with the Abel family of viewing time measures, as previously discussed. Perhaps the most significant way in which the Affinity measure of viewing time addresses concerns with the Abel is use of raw data. Specifically, though the Affinity has not yet been as thoroughly researched as the Abel measures, the raw data is immediately available to researchers and clinicians. Consequently, critical information is available for interpretation for both clinician and researcher use. Additionally, research has already begun with collecting "normal" samples of data so the Affinity may eventually become norm-referenced. Further research is underway in comparing Affinity results between "normal" samples and a variety of sexual offenders.

Association between PPG and Viewing Time Measures

Letourneau (2002) compared PPG and an Abel viewing time measures (the specific, commercially-available measure is unclear) in the only study of Abel viewing time measures performed outside his lab. Subjects were 57 adult males in a maximum security military prison (which served all four branches of the military). Subjects included enlisted military with sentences of five years or greater and officers. Of the 57 subjects, six volunteers denied any sexual offenses, including the index offense. The same subjects took both the PPG and viewing time measure. PPG stimuli included 16

audiotapes by the Association for Treatment of Sexual Abusers (ATSA). Tapes were three minutes in length and described adult masturbation or sexual interaction between an adult male and another person. There were four tapes in four categories: adult female consenting, adult male consenting, minor female compliant or coercive, and minor male compliant and coercive. The author reported there was also a female adult rape tape, which does not appear to be available from ATSA Audiotapes. Additionally, for the child stimuli, no specific age cues were provided, subjects were instructed to “think of a child just the age that you like.” After the stimulus presentation was complete, subjects were asked what age child they thought about. The researcher used the maximum circumferential change that occurred during stimulus presentation or 30 seconds thereafter and then averaged the changes within each category. For the viewing time measure, subjects were presented with the full 22 stimulus categories and seven slides per category, but only 18 categories were assessed for reliability and validity. The researcher averaged the seconds per slide in each category and also trimmed the data by eliminating the highest outlier per category as the measure’s author does.

Letourneau (2002) found an internal consistency of 0.82 or higher on the PPG. Untrimmed viewing time internal consistency ranged from 0.60 to 0.90 and trimmed data was at 0.70 or above, with the exception of males two to four years of age. Unsurprisingly, the coefficient alpha improved with the trimmed data, consistent with previous critiques. Letourneau (2002) separated offenses into four categories: against young girls, against young boys, against adolescent girls, and rapists of adult women. Each of these groups was comprised of all men who had any victim in the category, so offenders could be in more than one category, which were then compared to all other

offenders. The researcher computed a correlation between PPG and viewing time results. PPG results were significantly associated with viewing time untrimmed data in the following three of four categories: female child (0.277), male child (0.607), and female adult rape (0.376). Letourneau also computed a point-biserial correlation between the victim and each measure. PPG and viewing time measures were both significantly correlated with male child victims; viewing time was also significantly associated with female adolescent victims. Interestingly, the PPG was significantly and negatively correlated with female child victims. To further test for validity in clinical use, Letourneau used cutoffs as would be done in clinical practice (any PPG response greater than 10% and the viewing time rule of thirds) to determine category of responses and calculate the hits and misses. PPG and viewing time were both significantly correlated with young boys; unexpectedly, PPG was significantly negatively associated with female children. The author later hypothesized that this unexpected finding was due to lack of specific cues regarding the age of the female children and no specific female adolescent stimulus for PPG. In comparison, the viewing time measure identified all but one offender as interested in adolescent females. Finally, the only category in which there was no significant association between viewing time and PPG results was the female adolescent category; again, this may be due to there not actually being a female adolescent category in the PPG.

Despite the seminal nature of Letourneau's (2002) study, there were several significant flaws. First, her sample size was quite small, as she had only 57 subjects. Letourneau's study also had problems with generalizability in several areas. First, the subjects were all military personnel and were all incarcerated for their offenses. Many

males who sexually offend are not incarcerated. Additionally, the nature of their occupation is highly specific and may not necessarily generalize to the overall population. Another concern with Letourneau's study is her use of an Abel viewing time measure, as the data obtained in this measure are not readily available to researchers. Consequently, verification and replication of the data is quite difficult. Letourneau utilized a measure of penile plethysmography that is not standardized, which is one of the significant critiques of the PPG overall. Additionally, the PPG stimuli she used were not specific to an age category, as subjects were only instructed to consider "just the age you like." This procedure does not provide the necessary stimulus control to assess for age preference. Another problem with Letourneau's study is her comparison of the particular offenders of interest with all the other offenders. Instead of using different offenders as a control group, controls should have been used to determine differences.

There has been only one study of the concordance between PPG and viewing time measures of sexual arousal and interest (Letourneau, 2002). Unfortunately, this study utilized a small and specific sample, unstandardized PPG, and an Abel viewing time measure, which has significant documented problems. The purpose of the current study is to examine the concordance between a PPG system and viewing time measure using a larger and more diverse sample, standardized PPG system (i.e, Monarch 21), and a better measure of viewing time (Affinity).

Improvements over Letourneau study

The current study offered an improvement over Letourneau (2002) in several areas. First, subjects in the present study came from a general population, some of whom remained in outpatient treatment, some of whom went on to be incarcerated, and others

who have been released from prison or jail settings. Consequently, results of the current study were likely to be more generalizable than Letourneau's study. Another advantage of the current study was the number of subjects included. As previously noted, Letourneau's study included a total of 57 subjects; the current study had approximately 100 subjects. Another advantage of the current study was the utilization of a PPG method that addresses issues related to "faking" which have been researched. A final advantage of the current study's PPG was utilization of a standardized method of penile plethysmography. That is, the core stimuli were consistent each time this PPG system was utilized, in comparison with another PPG system that is not packaged with any stimuli. Consequently, the current PPG system addresses several concerns researchers have expressed with PPG. Additionally, the current study utilized the Affinity viewing time measure so data gathered were readily available to the researchers. Finally, norm-referenced research has begun with the Affinity, thus further enhancing the utility of this measure and applicability to a variety of subjects.

The current study examined the relationship between the results of the Affinity (a measure of viewing time) and a standardized PPG system (the Monarch 21). It was hypothesized the two measures will highly correlate in regard to key variables of sexual interest and sexual arousal by age groups and gender of the stimulus materials.

Statement of the Problem

Penile plethysmography is the single best predictor of sexual recidivism in those who have sexually offended and has an extensive body of literature regarding supporting its validity when used with adult males who have sexually offended. However, there are many concerns and problems with PPG including no standardization of administration,

no standardized stimuli, lack of standardized training, and dissimulation attempts. The aforementioned concerns are addressed via the Monarch 21 PPG system, as there are one set of standardized stimuli, training is consistent and has been conducted by the same individual for many years, technicians have a script regarding their client interactions and administration procedures, and attention and processing tasks are included to detect and discourage dissimulation. Consequently, the Monarch 21 PPG system is the best available PPG system.

Despite the significant improvement that Monarch 21 has made towards addressing concerns regarding PPG, several issues remain. First, extensive training is involved in learning to administer and interpret PPG. Additionally, the PPG testing process is both time-consuming (between two and three hours) and intrusive to the client. PPG may not be utilized with females and is of questionable ethics to administer to adolescent males. Another barrier to PPG testing is the significant financial investment required to obtain, utilize, and train in a system. Measures of viewing time address these concerns. That is, the Affinity viewing time measure is inexpensive to purchase, requires little training for administration, is not intrusive, and is quickly administered (approximately 15 minutes). If results of the Affinity may be shown to be positively associated with PPG results, this measure may be useful as a replacement or adjunct to PPG testing. This may assist in more effective and efficient assessment of those who sexually offend. Additionally, if the Affinity is positively associated with PPG, Affinity may be evaluated in the future for use with adolescents and females.

Research Questions

1. What is the association between the 18 (17 presented) core stimuli of Monarch 21 PPG and the eight categories in Affinity? This will compare the measures as they are used clinically (i.e., Monarch individual stimuli and Affinity categories).
2. Is the association between the coercive Monarch 21 stimuli and the Affinity significantly higher than the association between the persuasive Monarch 21 stimuli?
3. What is the association between eight collapsed Monarch 21 PPG stimuli and the eight categories in Affinity?
4. What is the association between Monarch 21 PPG and Affinity sexual deviance difference scores?
5. What is the association between Monarch 21 PPG and Affinity sexual deviance ratios?

Hypotheses

It was hypothesized that:

1. There will be a significant association between each individual Monarch 21 stimulus and the analogous (based upon age and gender) Affinity category. This hypothesis involves 18 comparisons, which may be seen in Table 1.*

Table 1

Correlations between Monarch 21 PPG and Affinity for Hypothesis 1

| Affinity category | Monarch 21 categories |
|---------------------------|--|
| Adult female (ADF) | Female challenge (FC) |
| Adult female (ADF) | Female adult normal/persuasive (FAN/P) |
| Adult female (ADF) | Female adult coercive (FAC) |
| Adult male (ADM) | Male challenge (MC) |
| Adult male (ADM) | Male adult persuasive (MAP) |
| Adult male (ADM) | Male adult coercive (MAC) |
| Juvenile female (JUF) | Female teen persuasive (FTP) |
| Juvenile female (JUF) | Female teen coercive (FTC) |
| Juvenile male (JUM) | Male teen persuasive (MTP) |
| Juvenile male (JUM) | Male teen coercive (MTC) |
| Pre-juvenile female (PJF) | Female grammar persuasive (FGP) |
| Pre-juvenile female (PFJ) | Female grammar coercive (FGC) |
| Pre-juvenile male (PJM) | Male grammar persuasive (MGP) |
| Pre-juvenile male (PFM) | Male grammar coercive (MGC) |
| Small child female (SCF) | Female pre-school persuasive (FPP) |
| Small child female (SCF) | Female pre-school coercive (FPC) |
| Small child male (SCM) | Male pre-school persuasive (MPP) |
| Small child male (SCM) | Male pre-school coercive (MPC) |

2. The associations for coercive stimuli will be significantly higher than the associations found for the persuasive stimuli, as coercive segments are more negative and thus may provide better discrimination. This hypothesis involves eight comparisons, which may be seen in Table 2.*

Table 2

Monarch 21 PPG Correlations with Affinity Categories in which Persuasive and Coercive Stimuli were Compared for Hypothesis 2

| Monarch 21 categories | | Affinity categories |
|------------------------------|----------------------------|---------------------|
| Persuasive | Coercive | |
| Female adult normal | Female adult coercive | Adult female |
| Male adult persuasive | Male adult coercive | Adult male |
| Female teen persuasive | Female teen coercive | Juvenile female |
| Male teen persuasive | Male teen coercive | Juvenile male |
| Female grammar persuasive | Female grammar coercive | Pre-juvenile female |
| Male grammar persuasive | Male grammar coercive | Pre-juvenile male |
| Female pre-school persuasive | Female pre-school coercive | Small child female |

3. There will be a significant positive association between the eight collapsed Monarch 21 stimuli and the eight categories in Affinity, which will be higher than any of the other above correlations. This higher correlation is hypothesized because there are more stimuli to comprise an individual scale. This hypothesis involves eight primary comparisons, which may be seen in Table 3 and 18 secondary comparisons which may be seen in Table 4.*

Table 3

Affinity Categories Compared to Monarch 21 Collapsed Categories

| Affinity category | Monarch 21 categories |
|---------------------|---|
| Adult female | Female challenge Female adult normal/persuasive Female adult coercive |
| Adult male | Male challenge Male adult persuasive Male adult coercive |
| Juvenile female | Female teen persuasive Female teen coercive |
| Juvenile male | Male teen persuasive Male teen coercive |
| Pre-juvenile female | Female grammar persuasive Female grammar coercive |
| Pre-juvenile male | Male grammar persuasive Male grammar coercive |
| Small child female | Female pre-school persuasive Female pre-school coercive |
| Small child male | Male pre-school persuasive Male pre-school coercive |

Table 4

Comparisons of Correlations between the 8 collapsed Monarch 21C categories and the 16 Individual Monarch 21 Stimuli

| Collapsed category | Monarch 21 individual stimulus |
|------------------------|--------------------------------|
| Female adult (FAD) | Female challenge |
| Female adult (FAD) | Female adult normal |
| Female adult (FAD) | Female adult coercive |
| Male adult (MAD) | Male challenge |
| Male adult (MAD) | Male adult persuasive |
| Male adult (MAD) | Male adult coercive |
| Female teen (FT) | Female teen persuasive |
| Female teen (FT) | Female teen coercive |
| Male teen (MT) | Male teen persuasive |
| Male teen (MT) | Male teen coercive |
| Female grammar (FG) | Female grammar persuasive |
| Female grammar (FG) | Female grammar coercive |
| Male grammar (MG) | Male grammar persuasive |
| Male grammar (MG) | Male grammar coercive |
| Female pre-school (FP) | Female pre-school persuasive |
| Female pre-school (FP) | Female pre-school coercive |
| Male pre-school (MP) | Male pre-school persuasive |
| Male pre-school (MP) | Male pre-school coercive |

4. There will be a significant positive association between the Monarch 21 PPG and Affinity sexual deviance differential scores. This is a primary hypothesis and will consist of only one correlation.
5. There will be a significant positive association between the Monarch 21 PPG and Affinity sexual deviance ratios. This is a primary hypothesis and will consist of only one correlation.

*This was an exploratory hypothesis that involved examining multiple correlations using the Bonferroni correction. In addition to testing for statistical significance, the magnitude of the effect size was examined to assess for clinical utility. As per Cohen (1988), correlations below 0.1 were considered to have no clinical utility. Additionally, the following effect sizes were utilized: a) $0.1 \leq r < 0.3$ is small; b) $0.3 \leq r < 0.5$ is moderate; c) $0.5 \leq r$ is large.

CHAPTER III

Methods

Design

A correlational design was used to examine the relationship between key variables in the Monarch 21 Penile Plethysmograph and the Affinity. These key variables included sexual deviance ratio, sexual deviance differential, and percent of response (erection throughout PPG or time viewing each category in Affinity).

Participants

Subjects consisted of 96 adult males who were referred to an outpatient psychological clinic for treatment of sexual offending behaviors, psychosexual evaluation, or sexual arousal testing as part of treatment. All testing for this study was performed as part of the evaluation process and was not additional to any testing for which subjects had entered the clinic. That is, the data utilized in this study were part of clinical records. As part of their intake paperwork, subjects were given a consent form for their data to be utilized in future research, divorced from all identifying information. Females were not included in this study, as they are not able to take the penile plethysmograph. Non-English speakers also were excluded as both measures are presented in English only. Finally, males under the age of 18 were excluded due to ethical concerns regarding their age.

Measures

Penile Plethysmograph

The Monarch 21 Penile Plethysmograph system was utilized for PPG. The Monarch 21 system is the only standardized PPG system available for public use. This

system is manufactured by Behavioral Technology, Inc. All testing was done utilizing a Barlow circumferential strain gauge or an indium gallium circumferential gauge. Due to concerns regarding comfort, all attempts were made for subjects to use the Barlow strain gauge. However, in certain males, accurate measurement is not possible with the Barlow strain gauge (e.g., extremely overweight males and sometimes uncircumcised males).

Stimuli

Core stimuli, which are utilized in all assessments, vary by age group (i.e., preschool, grammar-school, adolescent, and adult), gender, and level of manipulation (persuasion versus coercion). In addition to the aforementioned 16 stimuli, two more stimuli are considered core stimuli (Female Challenge and Male Challenge), one of which is always presented during testing (depending upon the subject's self-reported sexual orientation). The original 16 core stimuli are presented for 130 seconds and consist of both audio and visual stimuli. The first five seconds are an image of the individual the same age and gender of the following audio, to "prime" the subject for the audio segment. The following audio segment lasts for 85 seconds and is narrated by a male, in the first person, describing sexual interaction with the person of interest (including age, gender, and level of manipulation). The final 40 seconds of stimulus material consists of four images of the age and gender of the person just described, lasting for 10 seconds each. Additionally, during stimulus presentation, subjects were asked to listen for tones during the audio and watch for stars during the video. When subjects noted a tone, they were to press a button once if they believe the interaction to be persuasive and twice if they believe the interaction to be coercive. Subjects were to press the button when they saw a star during the presentation of video. In this fashion, subjects were required to

actively monitor and cognitively process the audio and video portions of the stimuli to deter dissimulation. All images in the aforementioned stimuli were clothed. The Male Challenge and Female Challenge segments are exclusively visual. These segments were presented to determine if subjects evidenced any arousal, even to strongly sexual stimuli. The Challenge segments show individuals in various stages of dress and were selected to match the subject's sexual orientation.

The first stimulus presented, though not examined for this study, was a neutral stimulus with an image of a landscape and a male narrating a peaceful scene. This was to ensure the equipment was properly functioning, subjects were able to see and hear stimuli, and understood instructions.

Setting

Subjects were seated in a comfortable, reclining chair that was covered with medical paper in a small room. Subjects were alone in the room throughout the evaluation process, and they wore a headset with earphones and microphone in order to communicate with the technician. Subjects had a respiration belt around their chests, a galvanic skin response sensor on one finger, and the strain gauge on the penis, halfway up the penile shaft. Stimuli were presented via the headset and video goggles.

Viewing Time Measure

The Affinity, available for purchase from David Glasgow, Ph.D., was utilized for this study (Glasgow, Osborne, and Croxen, 2003). The Affinity is a measure of viewing time, including stimuli of varied ages (small child, pre-juvenile, juvenile, and adult) and genders, consisting of eight categories. This measure was originally designed for individuals with learning disabilities, but appears to have broader application. Subjects

were informed this is a test of sexual interest. First, subjects saw a screen with eight drawings of prototypical human shapes in which size and ratio of body parts identify different gender and age categories. Two buttons are under these images, “Most Attractive” and “None.” Subjects were instructed to select (via mouse) which person they found most sexually attractive. That image was then highlighted and when the subject pressed the “Most Attractive” button, the image disappeared. This procedure was followed until the subject indicated none of the remaining images were sexually attractive, at which point he pressed the “None” button. Once the “None” button was pressed, both buttons were replaced with a single button “Most Unattractive.” The subject then was instructed to select a person who would be most sexually unattractive. Again, the subject selected the prototypical image with a mouse and, when certain the selection was correct, pressed the “Most Unattractive” button. This procedure continued until no more images remain.

Once the ranking task is completed, subjects began a rating task. It is during the rating task that viewing time was covertly measured. The first image presented was a “practice” image and was not included in data analysis. The screen showed an image, a scale below that had eight areas labeled unattractive, one neutral, eight areas labeled attractive, and a “Next” button. The area over which the subject had his mouse was highlighted, so he could see which area he is choosing. The time between initial stimulus presentation and rating choice (when subject clicks on rating area) was the measured viewing time. Post-task latency was the time between rating selection and pressing the “Next” button.

Stimuli consisted of seven pictures in each of eight categories. All images were of fully clothed individuals. In initial stimulus selection, judges were presented with many pictures and were asked to determine which category each image represented. If all five judges did not agree upon the stimulus category, the image was removed. Thus, all stimuli were presumed to accurately represent the age and gender of the category to which they belong. This process resulted in seven images for each category (Glasgow, Osborne, and Croxen, 2003). Results also showed that post-task latency is unstable, but on-task latency (or the viewing time) is stable. Thus, post-task latency was not used.

Procedure

Subjects entered an outpatient clinic for treatment, sexual arousal testing, or psychosexual evaluation. Sexual arousal testing was part of the intake procedure for treatment at this clinic. Upon arriving for their appointment, subjects were presented with intake paperwork requesting demographic information and regarding testing. As part of the testing packet, a consent form was included in which clients were asked if their raw data (divorced from identifying information) may be utilized for future research. Clients were able to ask questions regarding this process and it was explained that their consent, or lack thereof, would not impact any further interactions they had as part of the evaluation or treatment process. Indeed, once this consent process is completed, the information was not further provided to evaluators.

Subjects began the evaluation process once they completed intake paperwork. Typically, the process was completed in two consecutive days in which both the Affinity and Monarch 21 PPG were administered. Both measures were utilized with standard instructions included in the training of these methods. The Monarch 21 PPG included a

script of the interactions between the technician and subject. Additionally, a verbatim script was utilized when explaining the Affinity.

Data for this study was gathered, collated, and archived as part of a larger study prior to the beginning of this study. Consequently, only archival data, divorced of all identifying information was used.

CHAPTER IV

Results

PPG data consisted of the maximum arousal during stimulus presentation less the baseline at the start of the stimulus [“maximum arousal”] for each of the 17 core stimuli to which the subjects were exposed. The maximum arousal was then transformed to percent of erection by dividing the maximum arousal by the total arousal evidenced in the 17 core stimuli. This transformation was done so there is a similar scale between the PPG and Affinity. Separate comparisons between either persuasive or coercive stimuli and Affinity data were also made. Additionally, Monarch 21 data from all 17 core stimuli (to which each subject was exposed) was collapsed into eight groups that are analogous to Affinity categories as in Table 3.

Affinity data includes ratings, viewing time, and post-task latency to each image. Viewing time is the variable that was examined. Viewing time for each image was summed for each image category (i.e., adult females, adult males, juvenile females, etc.) for an overall category score. These scores were then transformed to percent of viewing time for each category. A correlational analysis was conducted between the various aforementioned PPG categories and these eight Affinity categories.

Sexual deviance ratios (highest adult divided by highest child) and differences scores (highest adult less highest child) were also computed for both PPG and Affinity results for each subject. These statistics were also correlated to examine the relationship between the PPG and Affinity.

*Analyses of Hypotheses**Hypothesis One*

It was hypothesized that there would be a significant association between each individual Monarch 21 core stimulus and the analogous Affinity category (based upon a cross between age and gender), resulting in 18 comparisons. A list of the 18 comparisons may be seen in Table 1. A two-tailed Pearson correlation between these variables was utilized, with a Bonferroni correction due to the multiple correlations. An alpha level of 0.003 was used to identify significant correlations, based upon the Bonferroni correction. Three of the 18 correlations were significant at the $p \leq 0.003$ level. The four significant associations were found between the following categories: the Female Grammar Persuasive and the Pre-Juvenile Female ($p = 0.002$, $N = 96$); the Male Adult Persuasive and the Adult Male ($p = 0.002$, $N = 96$); and the Male Teen Coercive and the Juvenile Male categories ($p = 0.003$, $N = 96$). Table 5 presents the correlations between the Monarch 21 and Affinity categories. The association between the Female Grammar Persuasive and Pre-Juvenile Female categories was moderate ($r = 0.30$, $N = 96$). The associations between the Adult Male categories (Male Adult Persuasive and ADM, $r = 0.31$, $N = 96$) and the Juvenile Male categories (Male Teen Coercive and JUM, $r = 0.30$, $N = 96$) were also moderate. It is noteworthy that though the correlation between the Monarch 21 Male Challenge segment and the Affinity Adult Male segment was quite high ($r = 0.57$, $N = 10$), this did not reach statistical significance due to the low number of members in the Male Challenge group.

Table 5

Correlations between Monarch 21 PPG and Affinity for Hypothesis 1

| Monarch 21 | Affinity | | | | | | | |
|------------|----------|-------|-------|-------|-------|------|------|-------|
| | ADF | ADM | JUF | JUM | PJF | PJM | SCF | SCM |
| FC | 0.16 | | | | | | | |
| FAN/FAP | 0.29* | | | | | | | |
| FAC | 0.03 | | | | | | | |
| MC | | 0.57 | | | | | | |
| MAP | | 0.31* | | | | | | |
| MAC | | 0.06 | | | | | | |
| FTP | | | -0.06 | | | | | |
| FTC | | | -0.02 | | | | | |
| MTP | | | | 0.09 | | | | |
| MTC | | | | 0.30* | | | | |
| FGP | | | | | 0.30* | | | |
| FGC | | | | | 0.16 | | | |
| MGP | | | | | | 0.10 | | |
| MGC | | | | | | 0.15 | | |
| FPP | | | | | | | 0.04 | |
| FPC | | | | | | | 0.03 | |
| MPP | | | | | | | | 0.05 |
| MPC | | | | | | | | -0.08 |

* $p < 0.01$.

Hypothesis Two

The second hypothesis was that the associations found between the Monarch 21 coercive stimuli and the analogous Affinity categories (based upon age and gender) would be higher than the associations found between the Monarch 21 persuasive categories and the analogous Affinity categories (again, based upon age and gender). This hypothesis resulted in eight comparisons of correlations. Fischer's z was used to compare the relative strength of these associations (Hays, 1981).

Significant differences in the correlations between the Monarch 21 stimuli and the analogous Affinity categories were found (though only one difference occurred in the hypothesized direction, with the other two being counter to the hypothesized direction). The significant differences in the correlations were found as follows: Female Adult Normal and Female Adult Coercive, Male Adult Persuasive and Male Adult Coercive,

and Male Teen Persuasive and Male Teen Coercive. All other associations were found to be being statistically nonsignificant. However, the coercive Monarch 21 stimuli had a significantly higher correlation with the Affinity categories than did the persuasive Monarch 21 stimuli in only one of the three significant comparisons: The Male Teen Coercive association was significantly higher than the Male Teen Persuasive association ($z = -2.16, p < 0.05$). In the other two significant comparisons (Adult Female [$z = 3.19, p < 0.05$] and Adult Male [$z = 2.56, p < 0.05$]), the persuasive segments had significantly higher associations with the corresponding Affinity categories, which was counter to the hypothesis.

Hypothesis Three

The third hypothesis was that there would be a significant positive association between Monarch 21 collapsed categories (in which stimuli were combined based upon age and gender) and the analogous Affinity categories and that the association found between the collapsed Monarch 21 categories and the Affinity categories would be significantly higher than the correlations found between individual Monarch 21 categories and their analogous Affinity categories. Thus, this hypothesis is two-fold. First, the Monarch 21 categories were collapsed into eight logically-derived variables based upon age and gender; these eight variables were used in a correlational analysis with Affinity categories. Next, the correlations between the collapsed Monarch 21 variables and the Affinity categories (eight correlations) were compared to the correlations between each individual Monarch 21 category and its' analogous Affinity category (18 correlations). First, however, Cronbach's Alpha was utilized to ensure internal consistency in the Monarch 21 collapsed categories, as these categories originally

were derived in a logical fashion, rather than empirically-validated. The Cronbach's Alphas for all but two of the collapsed scales were greater than or equal to 0.75 (see Table 6). The two scales with lower consistency were Male Grammar School (0.47), comprised of Male Grammar Coercive and Male Grammar Persuasive, and Male Preschool (0.42) which was comprised of Male Preschool Persuasive and Male Preschool Coercive stimuli. Thus, six of the eight collapsed categories had acceptable internal consistency, for the comparisons as hypothesized.

The next part of this hypothesis required a correlational analysis to be performed between the collapsed categories and the Affinity categories so the strength of the associations could be compared. Correlations between the collapsed Monarch 21 categories and the analogous Affinity categories are reported in Table 7. Five associations were significant, four of which were significant at the $p < 0.01$ level. The significant associations were found for female adults, male adults, teenage males, grammar-school females, and grammar-school males.

Table 6

Cronbach Alphas for Collapsed Monarch 21 Categories

| Monarch 21 categories | Cronbach Alpha |
|--------------------------------|----------------|
| Female challenge | 0.75 |
| Female adult normal/persuasive | |
| Female adult coercive | |
| Male challenge | 0.78 |
| Male adult persuasive | |
| Male adult coercive | |
| Female teen persuasive | 0.78 |
| Female teen coercive | |
| Male teen persuasive | 0.85 |
| Male teen coercive | |
| Female grammar persuasive | 0.84 |
| Female grammar coercive | |
| Male grammar persuasive | 0.47 |
| Male grammar coercive | |
| Female pre-school persuasive | 0.83 |
| Female pre-school coercive | |
| Male pre-school persuasive | 0.42 |
| Male pre-school coercive | |

Table 7

Correlations between Monarch 21 Collapsed Categories and Affinity Categories

| Monarch 21 | Affinity | | | | | | | |
|------------------|----------|--------|-------|--------|--------|-------|------|------|
| | ADF | ADM | JUF | JUM | PJF | PJM | SCF | SCM |
| Adult female | 0.32** | | | | | | | |
| Adult male | | 0.52** | | | | | | |
| Teenage female | | | -0.08 | | | | | |
| Teenage male | | | | 0.28** | | | | |
| Grammar female | | | | | 0.33** | | | |
| Grammar male | | | | | | 0.25* | | |
| Preschool female | | | | | | | 0.14 | |
| Preschool male | | | | | | | | 0.08 |

* $p < 0.05$, ** $p < 0.01$.

Fisher's z test was performed to compare the strength of the associations found between the Monarch 21 collapsed categories and the Affinity categories and the correlations found between the individual variables and the Affinity categories. Results of the Fisher's z test may be found in Table 8. Of the 18 comparisons between the collapsed correlations and the individual categories, three were found to be significant, in the expected direction. Specifically, the following associations were found to be significant: the Monarch 21 Adult Female collapsed association with Affinity adult female ($r = 0.32$) was significantly larger than Monarch 21 Female Adult Coercive with

Affinity adult female [($r = -0.02$) and $z = 3.35$, $p \leq 0.05$]; the Monarch 21 Adult Male collapsed association with Affinity adult male ($r = 0.52$) was significantly larger than the association between Monarch 21 Male Adult Persuasive and Affinity adult male [($r = 0.32$) and $z = 1.97$, $p \leq 0.05$]; the Monarch 21 Adult Male collapsed association with Affinity adult male ($r = 0.52$) was significantly greater than the association between the Monarch 21 Male Adult Coercive and the Affinity male adult [($r = 0.02$) and $z = 4.81$, $p \leq 0.05$].

Additionally, the larger correlation between collapsed Monarch 21 categories and Affinity categories trended in the predicted direction for the following 10 Monarch 21 categories (with the collapsed categories having a larger association with same-age Affinity categories than these individual Monarch 21 categories): Adult Female Persuasive ($r = 0.28$; $z = 0.41$), Female Challenge ($r = 0.16$; $z = 1.07$), Male Teen Persuasive ($r = 0.15$; $z = 1.28$), Female Grammar Coercive ($r = 0.20$; $z = 1.23$), Male Grammar Persuasive ($r = 0.19$; $z = 0.62$), Male Grammar Coercive ($r = 0.20$; $z = 0.53$), Female Preschool Persuasive ($r = 0.11$; $z = 0.34$), Female Preschool Coercive ($r = 0.10$; $z = 0.36$), Male Preschool Persuasive ($r = 0.19$; $z = 0.62$), and Male Preschool Coercive ($r = 0.20$; $z = 0.53$).

Hypothesis Four

Pearson's correlation was performed between the sex deviance differential scores found on the Monarch 21 PPG and the Affinity measure of viewing time. A significant association was found ($r = 0.20$, $p = 0.05$), though the effect size was small as the correlation was less than 0.3 (Cohen, 1988).

Hypothesis Five

Pearson's correlation was performed between the sex deviance ratio scores found on the Monarch 21 PPG and the Affinity measure of viewing time. A significant association was found ($r = 0.23, p=.03$), though the effect size was small as the correlation was less than 0.3 (Cohen, 1988).

CHAPTER V

Discussion

The goal of the current study was to evaluate if there is a positive association between the results of two measures utilized in the evaluation of males who have sexually offended: the Monarch 21 PPG and the Affinity measure of viewing time. This association was evaluated by performing correlational analyses of several key variables of both measures. Overall, the associations found between the Monarch 21 PPG and the Affinity suggest that these measures are at least somewhat comparable and further research should be done regarding the Affinity, its association with the PPG, and the predictive validity of the Affinity for sexual recidivism. Findings related to each hypothesis are discussed below, as well as the strengths and limitations of this study and recommendations for future research.

Hypothesis One

Statistically significant similarities were found between four of the 18 individual Monarch 21 categories and Affinity categories, though 10 more categories were trending toward, but did not reach significance. As it was hypothesized that the collapsed categories would have greater association with the Affinity categories due to the increased number of “test items” in the collapsed (versus the individual) categories, it was expected that there would be few individual categories which would achieve the power necessary to reach statistical significance. That is, these results may be an artifact of having few samples of behavior (e.g., inadequate power). Additionally, however, it is noteworthy that the effect size for the association between the Male Challenge segment and the analogous Affinity category was high ($r = 0.57$), though it did not reach statistical

significance ($p = 0.09$). It should be noted that this association was only made for the 10 subjects who were shown the Monarch 21 Male Challenge segment; consequently, this lack of significant association appears to be an artifact of the lack of power for this particular association. Consequently, future research regarding the association between the Monarch 21 PPG and the Affinity would benefit from a larger sample of individuals who have been shown the Male Challenge segment.

It is unclear in which way the direction and magnitude of the association would change if those shown the Male Challenge segment on the PPG also included males who self-identify as heterosexual. This may be a critical methodological difference between the PPG and Affinity since all individuals are shown the adult male segment on the Affinity, whereas only males who identify as homosexual are exposed to the Male Challenge segment on the Monarch 21 PPG. It is also possible that the high magnitude of the association is an artifact of the relatively small, homogenous group shown the Male Challenge segment on the Monarch 21 PPG.

Hypothesis Two

Based upon previous literature which suggested coercive stimuli are effective at discriminating between offenders and non-offenders (Quinsey and Chaplin, 1988; Chaplin, Rice, and Harris, 1995; Barsetti, Earls, Lalumiere and Belanger, 1998), it was hypothesized that the strength of the association between the coercive stimuli on the Monarch 21 PPG and the Affinity would be stronger than the strength of the association between the Monarch 21 PPG persuasive stimuli and the Affinity. This was based on the underlying assumption that the Monarch 21 PPG, and specifically the presumed best Monarch 21 PPG predictor of recidivism, would be most strongly associated with the

Affinity. However, this was not found to be the case in this study as only three significant associations were found and only one was in the hypothesized direction.

Several possible explanations exist for this unexpected finding. First, it is possible that this association does exist, but this particular study did not have the power to detect it. However, this does not appear to be the case, as two of the comparisons were stronger in the unexpected direction. It is also possible that the Monarch 21 coercive stimuli are not violent or disturbing enough to have discriminated between offender and control groups. This theory is a distinct possibility, as the stimuli used in the Chaplin, Rice and Harris (1995) study was extremely disturbing and quite graphic, and described as traumatizing, which is in contrast with the Monarch 21 stimuli which, although troubling, are not nearly as graphic as the aforementioned stimuli. A final possibility is that the Affinity measure of viewing time is simply not associated with the portion of the Monarch 21 PPG that discriminates best between offender and control groups, suggesting it may not be a good predictor of sexual recidivism. This final hypothesis should be more thoroughly examined by including the Affinity in future studies related to the best predictors of sexual recidivism, as the current study does not predict recidivism, but instead examines different measures that may be useful in the future to predict recidivism, with the Affinity being a possible alternative to the PPG.

Hypothesis Three

When examining the Monarch 21 PPG more closely, it was hypothesized that combining categories by age into eight general categories, instead of maintaining 18 distinct stimuli, would result in a higher association with Affinity categories. This is because it is thought that the more similar items comprising a scale, the more internal

consistency the scale will have, resulting in a stronger association with the Affinity.

However, the consistency of the collapsed scales needed to be examined to determine if they truly did comprise one, coherent scale. Acceptable internal consistency was found for 6 of the 8 categories and marginally acceptable internal consistency was found for the remaining 2 categories. Consequently, these eight scales were correlated with the eight Affinity scales and were found to have significant associations. However, the hypothesis was not only that the collapsed categories would be significantly associated with the eight Affinity categories, but also that these collapsed category associations with the Affinity would be significantly greater than the associations found between the Affinity and the individual categories.

As was expected, there were significant associations found between the eight collapsed categories and the Affinity categories. However, only three of these associations were *significantly* larger than the associations than between the 18 individual categories and the Affinity, though they were somewhat larger than 10 of the individual categories. It is possible that these associations were not found to be significantly larger than the associations for the individual categories due to lack of power, as a result not having enough subjects. Another possibility is that the collapsed categories are not more useful than simply considering the individual categories separately. However, as there is no more effort or stimulus exposure for the clients when utilizing collapsed categories, and there is a stronger association for 13 of the 18 categories, this approach may prove useful clinically. Nonetheless, this stronger association suggests the Affinity is more similar to the Monarch 21 PPG when Monarch 21 PPG stimuli are exclusively categorized by age and gender, rather than also including the nature of the interactions.

This is expected, as the Affinity does not include any depictions of interactions between individuals, which is dissimilar to the stimuli presented in the Monarch 21 PPG which narrates interactions between victim and offender.

Hypotheses Four and Five

Many researchers utilize methods to consider responses in context of each other, rather than utilizing absolute response values. Specifically, researchers review information such as the sex deviance differential or the sex deviance ratio. Consequently, these measures were computed for both measures and the associations were examined. The sex deviance differentials and ratios for the Monarch 21 PPG and the Affinity were found to be significantly associated, though the effect size was small. Again, these results suggest the measures were somewhat comparable. Though researchers have generally used the sex deviance differential and ratios to consider PPG results in context of each other, another method may be even more useful. Specifically, viewing the overall *shape* of the arousal pattern may yield further information that is lost in calculating the aforementioned statistics.

Barbaree and Marshall (1989) found five different arousal patterns in groups of offender and control subjects. It may be useful to compare the overall shape of the arousal and interest patterns found in the PPG and Affinity, respectively, rather than simply reducing the measures to two discrete ratio measures (deviance differential and deviance ratio). Indeed, if the pattern of responses were examined, the measures might be found to be more similar than was found in the current study. Comparison of response patterns may prove to be a useful direction for future research comparing the Affinity and Monarch 21 systems.

Strengths and Limitations

Overall, the associations found between the Monarch 21 PPG and the Affinity suggest that these measures are significantly associated, though they lack the strength of association necessary to presume psychometric equivalency. Though the current study furthered the field of research in this area by utilizing a larger sample size, having a sample that was more generalized than previous studies, and comparing more measures of PPG and viewing time, there remained several limitations. First, the general population from which this study draws is a fairly homogenous population - predominantly from Utah. Indeed, the population of Utah was estimated to be 93.8% Caucasian in 2005 (U.S. Census Bureau, n.d.) and generally people from this region hold the same religious beliefs. It would be useful to draw subjects from a sample more similar to the general population in this country to be more representative and generalizable. Additionally, though this sample is the largest of its kind, it remains relatively small and many of the comparisons, especially as related to the collapsed categories (as previously discussed) may evidence stronger associations with a larger sample size, thus furthering our understanding of the relationship between Monarch 21 individual stimuli and Affinity categories.

Finally, it is noteworthy that only a very small portion of the sample (10 subjects) identified as homosexual and thus were presented with the Male Challenge segment (as compared with the Female Challenge segment) on the Monarch 21 PPG. It would likely be more beneficial in the future to present all individuals with the Male Challenge segment, regardless of self-reported sexual orientation. This would be beneficial for several reasons. First, this would assist in the PPG being more analogous to the Affinity

and other viewing time measures. Another benefit would be to increase the number of subjects who were exposed to this stimulus, thus increasing the power of the statistical analysis. Finally, it would be useful to compare Monarch 21 PPG and Affinity arousal and interest patterns of individuals of a variety of sexual orientations as a whole, rather than grouping them into specific sexual orientation categories (e.g., homosexual, heterosexual).

Future Research

As previously noted, future researchers should address the limitations of the current study by utilizing a larger sample, a more heterogenous population, and exposing all study participants to the Monarch 21 Male Challenge segment. Additionally, current literature highlights the potential importance of viewing sexual arousal as measured on the PPG in context, rather than simply viewing each outcome measure in isolation. However, very little research has viewed PPG data in the context of overall patterns, rather than simply reducing PPG data to a series of discrete variables, thus losing potentially valuable data. Though a significant association was found between the Monarch 21 PPG and Affinity sexual deviance ratios and sexual deviance differentials, these measures might be found to be even more similar if examining the *shape* of the arousal pattern. Consequently, future researchers are encouraged to compare complex arousal pattern shapes, rather than focusing only on deviance differentials and ratios. This examination of complex pattern shapes should also be considered when studying which offenders have recidivated.

References

- Abel, G. G., Huffman, J., Warberg, B. & 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*, 81-95.
- Abel, G. G., Jordan, A., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the Abel Assessment for sexual interest. *Child Abuse & Neglect, 25*, 703-718.
- Abel, G. G., Lawry, S. S., Karlstrom, E., Osborn, C. A., & Gillespie, C. F. (1994). Screening tests for pedophilia. *Criminal Justice and Behavior, 21*, 115-131.
- Adams, H. E., Motsinger, P., McAnulty, R. D., & Moore, A. L. (1992). Voluntary control of penile tumescence among homosexual and heterosexual subjects. *Archives of Sexual Behavior, 21*, 17-31.
- Barbaree, H. E. & Marshall, W. L. (1989). Erectile responses among heterosexual child molesters, father-daughter incest offenders, and matched non-offenders: Five distinct age preference profiles. *Canadian Journal of Behavioural Science, 21*, 70-82.
- Barsetti, I., Earls, C. M., Lalumiere, M. L., & Belanger, N. (1998). The differentiation of intrafamilial and extrafamilial heterosexual child molesters. *Journal of Interpersonal Violence, 13*, 275 - 286.
- Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13*, 118-126.
- Chaplin, T. C., Rice, M. E., & Harris, G. T. (1995). Salient Victim Suffering and the Sexual Responses of Child Molesters. *Journal of Consulting and Clinical Psychology, 63*, 249-255.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Fischer, L., & Smith, G. (1999). Statistical adequacy of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11*, 195-205.
- Frenzel, R. R. & Lang, R. A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child sexual abusers. *Annals of Sex Research, 2*, 255-275.
- Freund, K. (1963). A laboratory method for diagnosing predominance of homo- or hetero-erotic interest in the male. *Behaviour Research and Therapy, 1*, 85-93.

- Freund, K. (1967a). Diagnosing homo- or heterosexuality and erotic age preference by means of a psychophysiological test. *Behaviour Research and Therapy*, 5, 209-228.
- Freund, K. (1967b). Erotic preference in pedophilia. *Behaviour Research and Therapy*, 5, 339-348.
- Freund, K. & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology*, 37, 100-105.
- Freund, K. & Watson, R. J. (1991). Assessment of the sensitivity and specificity of a phallometric test: An update of phallometric diagnosis of pedophilia. *Psychological Assessment*, 3, 254-260.
- Glasgow, D. V., & Croxen, J. D. (2003). *Viewing Time as a Measure of Paedophile Sexual Interest: Empirical Evidence, the Potential of Social Cognitions as a Theoretical Model, and Some Possible Implications for Assessment and Intervention*. Unpublished manuscript.
- Glasgow, D. V., Osborne, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: An application of single case methodology. *British Journal of Learning Disabilities*, 31, 96-102.
- Grossman, L. S., Cavanaugh, J. L., & Haywood, T. W. (1992). Deviant sexual responsiveness on penile plethysmography using visual stimuli: Alleged child molesters vs. normal control subjects. *Journal of Nervous and Mental Disease*, 180, 207-208.
- Hall, G. C., Proctor, W. C., & Nelson, G. M. (1988). Validity of physiological measures of pedophilic sexual arousal in a sexual offender population. *Journal of Consulting and Clinical Psychology*, 56, 118-122.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.
- Hanson, R. K. & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73, 1154-1163.
- 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.

- Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment, 4*, 502-511.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehard and Winston.
- Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment, 7*, 9-24.
- Kalmus, E., & Beech, A.R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*, 193-217.
- Kaufman, K. L., Rogers, D., & Daleiden, E. (1998). *Sexual interest in incarcerated adolescent offenders: A test-retest analysis*. Poster session presented at the annual meeting for the Association for Treatment of Sexual Abusers, Vancouver, BC.
- Laws, D. R., Hanson, R. K., Osborn, C. A., & Greenbaum, P. E. (2000). Classification of child molesters by plethysmographic assessment of sexual arousal and a self-report measure of sexual preference. *Journal of Interpersonal Violence, 15*, 1297-1312.
- Letourneau, E. J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse: A Journal of Research and Treatment, 14*, 207-223.
- Love, R. E., Sloan, L. B., & 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.
- MacCulloch, M. J. & Sambrooks, J.E. (1974). Sexual interest latencies in aversion therapy: A preliminary report. *Archives of Sexual Behavior, 3*, 289-299.
- 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.
- Marshall, W. L., Barbaree, H. E., & 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.
- Marshall, W. L. & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review, 20*, 807-822.
- McAnulty, R. D., & Adams, H. E. (1991). Voluntary control of penile tumescence: Effects of an incentive and a signal detection task. *The Journal of Sex Research, 28*, 557-577.

- Murphy, W. D., Haynes, M. R., Coleman, E. M., & Flanagan, B. (1985). Sexual responding of "nonrapists" to aggressive sexual themes: Normative data. *Journal of Psychopathology and Behavioral Assessment*, 7, 37-47.
- O'Donohue, W., & Letourneau, E. (1992). The psychometric properties of the penile tumescence assessment of child molesters. *Journal of Psychopathology and Behavioral Assessment*, 14, 123-175.
- Quinsey, V. L. & Chaplin, T. C. (1988). Preventing faking in phallometric assessments of sexual preference. *Annals of the New York Academy of Sciences*, 528, 49-58.
- Quinsey, V. L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology*, 17, 341-354.
- Quinsey, V. L., & Laws, D.R. (1990). Validity of physiological measures of pedophilic sexual arousal in a sexual offender populations: A critique of Hall, Proctor, and Nelson. *Journal of Consulting and Clinical Psychology*, 58, 886-888.
- Rosenzweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine*, 4, 150-157.
- Smith, G., & Fischer, L. (1999). Assessment of juvenile sex offenders: Reliability and validity of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A journal of Research and Treatment*, 11, 207-216.
- U.S. Census Bureau (n.d.). *State and County QuickFacts*. Retrieved September 23, 2007, from <http://quickfacts.census.gov/qfd/states/49000.html>
- Wright, L. W. & Adams, H. E. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment*, 16, 221-231.
- Zabarauckas, C. L. (2001). *Visual reaction time and sexual interest in male sexual offenders: moving away from penile plethysmography*. Unpublished master's thesis, University of Victoria, Victoria, British Columbia, Canada.
- Zamansky, H. S. (1956). A technique for assessment of homosexual tendencies. *Journal of Personality*, 24, 436-448.