2008-07-11

Non-Pedophilic Heterosexual Male Response to Affinity 2.0

Charles D. Crosby
Brigham Young University - Provo

NON-PEDOPHILIC HETEROSEXUAL MALE RESPONSE TO THE AFFINITY 2.0

by

Daniel Crosby

A dissertation submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Counseling Psychology and Special Education

Brigham Young University

March 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Daniel Crosby

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Lane Fischer, Chair
Date	Aaron Jackson
Date	Steve A. Smith
Date	Ellie Young
Date	Melissa Heath

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Daniel Crosby in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

_____ _____

Date

Lane Fischer
Chair, Graduate Committee

Accepted for the Department

_____

Aaron Jackson
Graduate Coordinator

Accepted for the College

_____

K. Richard Young
Dean, David O. McKay School of Education

ABSTRACT

NON-PEDOPHILIC HETEROSEXUAL MALE RESPONSE TO THE AFFINITY 2.0

Daniel Crosby

Department of Counseling Psychology and Special Education

Doctor of Philosophy

Given that the Affinity 2.0 is a largely untested instrument that purports to measure sexual attraction, it was unknown whether or not the instrument is temporally stable and would provide a characteristic response curve for normal heterosexual male sexual interest. This study examined the Affinity 2.0, a newly standardized viewing time (VT) instrument that purports to measure sexual interest. More specifically, the Affinity 2.0 was examined in regard to its temporal stability and ability to generate a characteristic curve for non-pedophilic, exclusively heterosexual male interests. This examination was carried out by administering the Affinity 2.0 twice on a sample of 88 self-reportedly non-pedophilic, exclusively heterosexual men. The results of this study find the Affinity 2.0 to be mildly to moderately temporally stable and capable of rendering a characteristic sexual interest curve. Subjects created a typical profile that showed longest viewing time to images of adult and juvenile females with low viewing times to all other categories of attractors. Implications for norm-referenced decision making are discussed.

ACKNOWLEDGMENTS

I would like to thank my wife Katrina whose love, support, and hard work have been instrumental in the completion of this dissertation. Her commitment to me and my work has given me strength to complete what was at times a very difficult undertaking. I am greatly indebted to her, and want to her to know that her contribution to this work was significant and much-appreciated.

I would also like to thank my family for their support throughout this project. My parents have been a great support to me throughout my education, and if I have been able to accomplish anything worthwhile as a student, it is only as a direct result of their generosity and kindness. I wish to thank my sisters, Meg and Sarah, for their friendship and love during my college years. My time at BYU has been enriched by their being a part of it. My brother Garrett also deserves recognition, for contributing a vast body of music that served as inspiration for the hard times. Finally, I wish to thank my Nana (Margaret Byrd), for encouragement, empathy, and Southern cooking that helped sustain me throughout graduate school.

Finally, I wish to thank my dissertation chair, Dr. Lane Fischer, for his tireless support, humor and encouragement. Without his gentle prodding and patience, this project would likely not have been seen through to fulfillment. He has been a great benefit to me, both as a colleague and a friend, and I wish to thank him for that.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Chapter 1

Introduction

Having a reliable instrument that provides an "accurate measure and classification of sexual arousal and preference [is] a prerequisite to adequate research and clinical activity" (Wright & Adams, 1994, p. 221). The discovery of a reliable instrument has a variety of potential uses, including tracking sexual interest throughout the lifespan, monitoring therapeutic progress, as well as a screening device for detection of deviancy. Currently, the four methods of measuring and classifying sexual arousal in males are the clinical interview, self-report measures, penile plethysmography, and viewing time (Quinsey, Rice, Grant, & Reid, 1993). While each of these measures may reveal important information regarding sexual interest, not all are appropriate or accurate measures of sexual interest in males.

Despite the widespread use of the clinical interview, it has specific limitations that make it an impoverished source for obtaining accurate information with regards to sexual interest. Perhaps the most glaring limitation of the clinical interview is its inability to prevent dissimulation (Marshall, 1996). The flaws of the clinical interview are exacerbated when dealing with a subject as private as sexual interest. Respondents are likely to show an aversion to being questioned about something as personal as their sexuality and may answer in a socially desirable way in an effort to appear "normal" (Quinsey et. al., 1993). The clinical interview may be too subjective and open to dissimulation to be used when dealing with the sensitive issue of sexual interest.

As is the case with the clinical interview, self-report measures of sexual interest have many strengths but are weakened by subjectivity and the potential for dissimulation.

As reported by Marshall (1996), self- report measures are highly transparent, leaving them vulnerable to those wishing to give misleading results. Quinsey et al. (1993) also found that a misunderstanding of questions, a desire to appear *normal* and a reluctance to speak about sexual matters adversely affect the efficacy of self-report measures.

Self-report methods are typically paper and pencil measures of how physically attracted someone is to the picture of a given individual. Numerous research studies show a strong relationship between physical attraction and sexual attraction (Berscheid, 1981; Patzer, 1985; Adams, 1977; Dion et al., 1972). Morse, Grusen, & Reis (1976) found that the relationship between physical attractiveness and sexual attractiveness was especially pronounced in males. Furthermore, a meta-analysis of the extant research also suggests that ratings of physical attractiveness (and thus, to a large degree sexual attractiveness) are stable from a very young age, and across cultures (Langlois & Roggman, 1990).

Despite the aforementioned strengths of self-report measures, flaws in this method make its value as a measure of sexual interest questionable. As with the clinical interview, self-report measures of subjects' sexual preference or quasi-sexual preference are necessarily somewhat ambiguous and can be strongly affected by social desirability (Quinsey, et al.). Dissimulation is another concern for any self-report measure, and is especially troublesome when working with an issue as sensitive as sexual interest. A third weakness of self-report measures of physical attractiveness is that it is often unclear what aspect of physical attractiveness is being focused on. There are personal and cultural standards for beauty, and it is often unclear to the participant how they are to rate the image being viewed (Quinsey, et. al.). Thus, despite its strengths, the methodological

inadequacies of self-report measures make it a less effective method for measuring a subject as sensitive as sexual interest.

A more objective, yet highly intrusive method for measuring sexual interest in males is the penile plethysmograph (an instrument that measures penile tumescence). Currently, penile plethysmography (or phallometry) is the most scientifically accepted of all methods for assessing sexual interest (Quinsey & Chaplin, 1988). Scientific acceptance notwithstanding, phallometry is highly invasive and requires sexually explicit material in order to generate a genital response. These requirements make the use of the penile plethysmograph with adolescents and children ethically questionable (Marshall, 1996). Additionally, Marshall and Fernandez have also questioned the psychometric soundness of phallometry, citing potential problems with standardization, temporal stability, criterion validity, data formats, and internal consistency (2000). Consequently, for reasons of ethicality, practicality, and psychometric soundness, alternative methods to phallometry should be explored.

A fourth method for assessing sexual interest that bears further exploration is measuring sustained visual attention. Studies of viewing time have proved effective in discriminating between homosexual and heterosexual males and females, child molesters, groups of people who experienced high and low guilt around sexual topics, and sexually interested and uninterested patients (Harris, Rice, Quinsey, & Chaplin, 1995; Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996; Wright & Adams, 1994; Quinsey et al., 1993; Love, Sloan & Schmidt, 1976; Rosenzweig, 1942). Additionally, all of the aforementioned studies found significant correlations between sexual preference and viewing time.

Additional advantages to viewing time as a measure of sexual interest are that it is non-intrusive, covert, and does not necessarily require sexually explicit materials. As such, it could potentially used with children, adolescents, and others groups inaccessible by an invasive measure such as penile plethysmography.

Two instruments that utilize viewing time as a measure of sexual interest currently exist. The first is the Abel Assessment for Sexual Interest (AASI). Research on the psychometrics of the AASI give some cause for concern regarding the instrument's data format, normative base, temporal stability, validity, and reliability (Fischer & Smith, 1999), deficits in test-retest reliability (Smith & Fischer, 1999 ; Kaufman, Rogers, & Daleiden, 1998) and the possibility for dissimulation (Gray, 1999). A further weakness of the Abel Assessment is that it uses ipsative z-scores; which consequently robs researchers of the opportunity to compare scores across individuals or groups. Finally, the use of the AASI as a tool for clinical and research purposes is further hampered by the fact that users of the instrument are not given access to the raw data results (Fischer & Smith, 1999). Given the aforementioned shortcomings of the AASI as a research and clinical tool, it is a poor choice for researchers wishing to answer questions regarding male sexual interest.

The Affinity 2.0 is a second instrument designed to assess sexual interest on the basis of sustained visual attention. Developed by David Glasgow, the Affinity 2.0 is a computer-based assessment of sexual interest that uses surreptitious measures of viewing time to create an individual profile of relative sexual interest by gender and age (Glasgow, Croxen, & Osborne, 2003). The Affinity 2.0 has the added strength of providing raw scores to researchers, a condition which allows them to create norm-

referenced data. Furthermore, the Affinity 2.0 is non-invasive and uses pictures of fully clothed models in non-suggestive poses; thereby making use with younger populations more ethical. In summary, the Affinity 2.0 is conducive to clinical and research purposes in that it provides raw data, is non-invasive, and ethically sound.

Statement of Problem

Given that the Affinity 2.0 is a newer instrument, we do not yet know what the non-pedophilic, exclusively heterosexual male response is to it or whether that response is temporally stable.

Statement of Purpose

The purpose of this study is to determine the typical heterosexual male response to the Affinity 2.0, as well as to determine whether that response is temporally stable.

Chapter 2

Review of the Literature

Having a reliable instrument that provides an "accurate measure and classification of sexual arousal and preference [is] a prerequisite to adequate research and clinical activity" (Wright & Adams, 1994, p.221). The four extant methods for measuring and classifying sexual arousal in males are the clinical interview, self-report measures, penile plethysmography, and viewing time (Quinsey et al., 1993). This review of literature will examine the relative strengths and weaknesses of each of these methods as applied to a non-pedophilic, exclusively heterosexual population of males.

Four Extant Methods for Measuring and Classifying Sexual Interest

Clinical Interview

The first of the four methods for assessing sexual interest to be examined is the clinical interview. While the clinical interview has some strengths, it has been greatly criticized on the basis that detecting dissimulation in an interview is extremely difficult (Marshall, 1996). This fact, combined with the inherent subjectivity involved in the interview process, make it a technique that can often yield misleading results. Due to the sensitive nature of the topic of sexual interest, many respondents may alter their responses in an effort to appear *normal* and may be reticent about verbally sharing any information about their sexual interest (Quinsey et al.). These factors may compromise the validity of the interview and may not lead to results that are an accurate reflection of the respondent's true sexual interest.

Thus, while some respondents may provide accurate information in clinical interviews, the opportunity for undetected distortion is sufficiently great as to make the

validity of the clinical interview suspect. The tendency to dissimulate is also of special concern when working with those who wish to appear "normal" or wish to hide some aspect of their sexual interest (e.g., pedophiles). While it is not without its virtues, the clinical interview is subject to dissimulation, and may produce misleading information due to confusion, embarrassment, or societal pressures.

Self-report Measures

Self-report is a second method by which sexual interest has been examined in the past. Self-report is typically a paper and pencil measure of how physically or sexually attracted someone is to the picture of a given individual. Since physical attractiveness is known to play an extremely important role in determining marital and dating partner choice (Berscheid, 1981; Patzer, 1985), it is assumed that those who are found to be physically attractive will also typically be found to be sexually attractive. There are myriad studies that point to the connection between physical attractiveness and sexual attractiveness (Adams, 1977; Berscheid, 1981; Dion et al., 1972; Morse, Grusen, & Reis; 1976). Adams' (1977) review of research on physical attractiveness found that it is related to sex appeal, as well as to feelings of love and emotional arousal. Berscheid (1981) undertook a review of existing evidence that showed that physical attractiveness has a profound effect on dating choice, especially in males. In their 1972 study, Dion et al. found that physically attractive men and women were rated higher in the areas of excitement and sexual warmth than were less attractive persons. Morse, Grusen, and Reis (1976) found that males, but not females, listed physical attractiveness and sex appeal as the most important considerations when assessing a potential partner of the opposite sex.

Thus, it seems that the research suggests a relationship between physical attractiveness and sexual attractiveness, especially when males are the population being studied .

Physical attractiveness has typically been measured primarily by simple ratings of the attractiveness of stimulus persons on a Likert scale (e.g., Berry & McArthur, 1985; Cunningham, 1986; Horvath, 1981; Mueser et al 1984). Two other self-report methods mentioned by Quinsey et al. (1993) are asking participants to rank order the stimuli according to the amount of *beauty* they see in each stimulus (Cross & Cross, 1971) or to have them rank order a set of stimuli according to physical attractiveness (Cavior & Lombardi, 1973; Korthase & Trenholme, 1982). Despite its subjective nature, self-report measures of physical attractiveness actually have fairly high levels of interrater reliability. Langlois and Roggman's 1990 meta-analysis of recent studies indicates that even children as young as three to six months spend more time looking at the faces of adults judged as attractive than they do at the faces of adults judged as unattractive. Even more impressively, the results were stable across cultures and independent of the ethnic diversity of the adult face. A 1973 study by Cavior and Lombardi further buttresses the assertion that even very young children are aware of what they find attractive. In this study, the researchers asked children aged five to eight to rate the physical attractiveness of photographs of 11, to 17-year-olds. The results showed that by age seven or eight children could make ratings that closely approximated those of the 11 to 17-year-olds. Similarly, Dion (1973) found that children as young as three could make reliable ratings of the physical attractiveness of facial photos of children their same age. Again, the ratings of the three-year-olds closely matched adult ratings of the same group of facial photos.

Another flaw of self-report methods is that it is often unclear which measure of physical attractiveness is being focused on. Depending upon the nature of the instructions given to raters, they can represent the respondents' understanding of cultural standards of beauty, their personal preferences, or some combination of both (Quinsey, et al.). Some of the aforementioned cases had the added caution that raters were to rate the pictures according to their own personal standards instead of by cultural standards of beauty. However, other studies made no mention of what criteria were to be used when making ratings. Given the aforementioned limitations, it is possible that the participants' ratings of the slides may not be an accurate reflection of their true attraction.

Self-report measures of physical attractiveness have many strengths when trying to determine the sexual interest of an individual. Numerous studies suggest that even very young children are able to accurately rate individuals as physically attractive or unattractive. Another strength is that this trend seems to hold across cultures. Despite their many strengths, self-report measures also possess some methodological inadequacies. Self-report measures of subjects' sexual preference or quasi-sexual preference are necessarily somewhat ambiguous and can be strongly affected by demand characteristics (Quinsey, et al.). Dissimulation is another concern for any kind of self-report, and especially one that measures something as sensitive as sexual interest. Especially when working with groups of people who wish to hide their true sexual interests (e.g., pedophiles), self-report measures are hopelessly transparent and responses are easily falsified. The above-mentioned considerations make self-report measures a poor measure of sexual interest for some research purposes.

Penile Plethysmopgraphy

Penile plethysmography (or phallometry) is a third method used for measuring sexual interest. Currently, phallometric assessment is the most scientifically accepted of all methods for assessing sexual interest in males (Quinsey & Chaplin, 1988). Penile plethysmography is based on the fact that penis volume changes in response to pictures that are more or less sexually attractive to the person being measured (Freund & Costell, 1970). Thus, the penis of normal adult heterosexual males become most tumescent when presented with pictures of adult females and shows decreasing levels of penile tumescence as the age of the female being viewed decreases. Conversely, normal adult homosexual males show peak tumescence when presented with pictures of adult males, and show a steady decrease in tumescence as the age of the person in the picture decreases (Freund, Langevin, Cibiri, & Zajac, 1973). Penile plethysmography has also shown itself useful in discriminating persons with histories of child molestation from non-offenders, as well as determining the sexual preferences of child molesters (Freund, 1967).

Despite its strengths, phallometry has a number of weaknesses that make it unsuitable for some research purposes. One major drawback of phallometric assessment is that it uses an invasive procedure and the presentation of pornographic material to engender genital response. As such, phallometry poses ethical issues for dealing with children, adolescents, and other populations who would oppose such invasive procedures (Marshall, 1996). Additionally, Marshall and Fernandez have questioned the psychometric soundness of phallometry, citing potential problems with standardization, temporal stability, criterion validity, data formats, and internal consistency (2000). While

it has shown some promise, phallometry is expensive, is highly invasive, is ethically suspect with certain large populations, and has some questionable psychometric properties. Given these shortcomings, other methods of objectively ascertaining sexual interest should be pursued.

Viewing Time

Viewing time (sustained visual attention) is a measure that has been used successfully in identifying both normal and deviant sexual interest. Rosenweig's 1942 study was the first to experiment with viewing time as a measure of sexual interest. He discovered that psychiatric patients who were rated as interested in sexual topics looked at sexual stimuli longer than males who were rated as less interested.

Zamansky (1956) was able to use viewing time of non-erotic stimuli to differentiate between groups of homosexuals and non-homosexuals. In the Zamansky study, 20 male participants identified as homosexual and 20 male participants identified as heterosexual were asked to look at a series of pictures. The male and female models were all fully clothed, and the series also included neutral pictures of scenery. The pictures were presented in pairs: Male/Female, Male/Neutral, Female/Neutral, and Neutral/Neutral. Participants were then instructed to examine each picture carefully and judge which one covered more area (all pictures of male or female models were the same size while the sizes of the neutral pictures were of varying sizes). Participants were scored by the number of seconds they spent looking at each picture in the pair. As hypothesized, the results showed that homosexual males did spend more time looking at pictures of males than they did looking at females, and that heterosexual males spent more time looking at pictures of females than males.

In 1994, Wright and Adams were able to differentiate between groups of homosexual males and females and heterosexual males and females through the use of viewing time. A total of 80 participants were involved in this study. Twenty individuals were assigned to one of four groups (homosexual male, homosexual female, heterosexual male, heterosexual female) based upon their sexual orientation. Each participant was shown 20 slides that had been divided into three categories: nude males, nude females, neutral scenes. On each of these slides, a white dot had been placed at a different location. The participants were to find the white dot as quickly as possible and indicate that they had located it by pressing a button. Results from this experiment showed that heterosexual males spent more time with nude female slides, homosexual males spent more time with nude male slides, heterosexual females spent more time with nude male slides, and homosexual females spent more time with nude female slides. Based on these results, researchers were able to differentiate between the groups based on their reaction times.

Harris, Rice, Quinsey, and Chaplin (1995) used viewing time to discriminate between child molesters and normal heterosexual males. Participants were asked to look at a total of 70 photographic slides consisting of seven categories: neutral landscapes, female children between the ages of five and eight, male children between the ages of five and eight, female pubescents, male pubescents, female adults, and male adults. Twenty of the slides were used as practice data and warm up which depicted the participants in each of the categories fully clothed. After the warm up, all non-neutral slides depicted one nude person with the genitals visible. Results indicated that viewing time did indeed discriminate between child molesters and non-child molesters. Harris et

al. concluded that viewing time showed promise as an unobtrusive measure of sexual interest that could differentiate between child molesters and non-offending males.

Quinsey et al. (1996) discovered differences between male and female sexual interests using viewing time. This study used nude photos of both males and females at varying ages from pre-pubescence to young adulthood. The four hypotheses tested by this investigation are as follows: 1) Males and females should look at pictures of young adults of the opposite sex longer than adults of the same sex regardless of age. They should spend the least time viewing photos of prepubescent individuals. 2) The correlation between viewing time and sexual attractiveness should be shorter for female participants. 3) Females should look at photographs of prepubescent males for less time than males looking at prepubescent males. 4) Males should look longer at adult females than female participants look at adult males. All of these hypotheses were based on principles of evolutionary psychology. The first three hypotheses were confirmed by the study. The fourth analysis was in the direction of prediction, but failed to attain statistical significance. It was the conclusion of the researchers in this study that viewing time could reflect sexual interest and that variations in previous studies may have clouded the picture.

A surreptitious measure of viewing time has been shown to be effective in differentiating between individuals based on their sexual interest. It has also been shown to differentiate between groups of child molesters and non-child molesters (Harris et. al., 1995), homosexual males and females from heterosexual males and females (Wright & Adams, 1994; Zamansky, 1956). While it seems reasonable to say that viewing time does have the potential to differentiate between people of various sexual orientations, different

measures of viewing time have their respective strengths and weaknesses. The following section will examine the two existing measures of viewing time, examine their strengths and weaknesses, as well as their suitability for general research purposes.

A valid measure of sexual interest that includes pictures of fully clothed models in non-suggestive poses would be of great worth to the scientific community. Pictures of fully clothed models will not be offensive to the general population and will also have greater applicability, as they will be able to be used even with minors. Currently, there are two such tests designed to measure sexual interest that use fully clothed models. The first is the Abel Assessment for Sexual Interest (AASI) and the second is the Affinity 2.0.

Two VT Instruments

AASI

Developed in 1994, the AASI was introduced as a non-intrusive assessment of sexual interest that utilized viewing time as its primary measure. The AASI is comprised of a questionnaire and a device designed to measure viewing time. The questionnaire involves a self report of sexual behavior (Fischer, 2000). The apparatus that measures viewing time consists of a slide carousel attached to a viewing screen connected to a laptop computer. The slides consist of 160 fully clothed individuals in non-sexual positions of varying ages, gender, and ethnicity. There are also six slides of fully clothed individuals that give a depiction of one of six different paraphilias: sadomasochism against females and males, exhibitionism, voyeurism, frotterism, and fetishism (Fischer, 2000). If proven valid, the AASI would be valuable as a large scale screening device for sexual offenders. Because the models are clothed, it could be used with women, men, and

adolescents. This would make the AASI much more amenable to use with these populations than the penile plethysmograph.

However, as of this time, the AASI appears only to show promise. Abel reports an internal consistency of .84 to .90 for the AASI as measured by Chronbach's Alpha (Krueger, Bradford, & Grahm, 1998). In fact, in 1998, Abel and his colleagues published a report stating that the evidence of validity and reliability of scores obtained by the Abel Assessment are comparable to those obtained by using penile plethysmography. However, in 1999, Fischer and Smith questioned Abel's methods for obtaining such a high internal consistency, as well as proposing several validity issues. Fischer and Smith also showed that the reliability and validity evidence from the scores obtained from the AASI were highly suspect when used with adolescent sex offenders (Smith & Fischer, 1999). Additionally, several other researchers have also questioned the validity and reliability evidence of the Abel Assessment as a screening device for sex offenders (Kaufman et al., 1998; Fischer, 2000; Gray, 1999). Despite the numerous studies that question the validity and reliability of the AASI, as of 1999, it was being used in two countries, 36 states, 8 state's judicial systems and by approximately 300 therapists (Fischer & Smith, 1999). In their 1999 publication, Fischer and Smith warned that there are dangers in using a non-validated instrument when making clinical or judicial decisions. Obviously, clinical decision- making will be compromised when using a non-validated instrument, and judicial rulings will be open to appeal and possible reversal.

Another major weakness of the AASI is in the interpretation of the data collected. According to Fischer and Smith (1999), ipsative scores can only show intraindividual variation. The interpretation of intraindividual variation can be enhanced if the

underlying mean and standard deviation from which the ipsative score was created is reported. Currently, the Abel Assessment for Sexual Interest presents the data in ipsative z-scores for viewing time and does not report the underlying raw score category means or the standard deviation of category means for each individual. In the absence of such information, ipsative scales become analogous to ordinal scales. Thusly limited, the researcher can make statements indicating that a participant possesses one attribute more than another; however, it is impossible to indicate how much more or less of one attribute that the participant possesses when compared with other participants in the study.

Unfortunately, Abel refuses to release the raw data to any user of the instrument (Fischer & Smith, 1999). Abel's failure to release the raw data associated with individual participants is one of the reasons why the AASI is currently being advised against in some courtrooms. Said one judge of this practice, (Amarillo, Texas Appellate Court, 2002)

> In short, of what the formulas applied by Abel consist, how they were derived, and whether they have ever been subjected to analysis or testing goes utterly unmentioned by Mack or anyone else. For all we know, they and their components could be mathematically based, founded upon indisputable empirical research, or simply the magic of young Harry Potter's mixing potions at the Hogwarts School of Magic. (p. 683)

Certainly, the failure to release raw score data or explain how scores are computed on the AASI severely hampers its effectiveness and application, both in the courtroom and as a research tool.

Affinity 2.0

The Affinity 2.0 is a computer program developed to investigate sexual interest (Glasgow, 2003). The assessment was created by David Glasgow and is currently in its second revision. While the original version was specifically designed as an assessment for sexual interest of males with learning disabilities, the current version is licensed to be used as a clinical assessment tool with learning disabled adult male offenders as well as non-learning disabled adult male offenders. Glasgow (2003) has also approved the Affinity 2.0 for use in research and evaluation purposes with juvenile male and female offenders.

One major strength of the Affinity 2.0 is that it offers to report all scores in their raw score form rather than having the scores converted into ipsative z-scores as with the AASI (Fischer & Smith, 1999). The major advantage of having raw score data is that it is available for both research and clinical purposes. Additionally, the scores obtained from the Affinity 2.0 can be made into norm-referenced scores that can be used for subsequent normative study. Commensurability allows comparisons across individuals as well as groups.

Given that the Affinity 2.0 is a newly standardized instrument, we do not yet know what the non-pedophilic, exclusively heterosexual male response is to this instrument or whether that response is temporally stable.

The purpose of this study is to determine the typical heterosexual male response to the Affinity 2.0, as well as to determine whether the response is temporally stable.

Chapter 3

Methods

Participants

Criteria for Inclusion

Participants consisted of males with a minimum age of 18. Since the purpose of the study was to examine the viewing time response of a group of non-pedophilic, exclusively heterosexual males, the researchers wished to screen out individuals whose sexual interest was not exclusively heterosexual as well as those who had a history of pedophilia. Homosexual interest and history of pedophilia were screened for by responses to a questionnaire that was administered to each participant subsequent to administration of the measure of sexual interest. Any participants that expressed homosexual interest or a history of pedophilia were fully compensated, but were not included as part of the experimental group.

Location of Recruitment

Participants selected for this study were sampled from the large population of undergraduate psychology classes at Brigham Young University. Individuals attending psychology classes from Brigham Young University were asked to participate via a short presentation by a researcher. The researchers explained that the purpose of this study was to test a new device that purports to measure sexual interest. Potential participants were informed that they would complete a short questionnaire and look at several still images of fully clothed models depicted in every-day life situations and rate images on their sexual attractiveness or sexual unattractiveness. They were also instructed that they would repeat the process a second time two to four weeks later. Compensation for all

participants' full participation was two free movie tickets. Participants from this pool were instrumental in providing a sufficiently robust sample for the purposes of this study.

Number of Participants

According to the Central Limit Theorem (McClave & Sincich, 2006), a curve approaches normality with a sample size of n=30. The greater the n, the more it will approximate the normal curve and be representative of the population at large. These things considered, the researchers originally decided on a sample size of 100 individuals. It was thought that a sample of this size was robust enough to make statistical inferences about the population at large, but not so large as to make the collection process unnecessarily time-consuming and expensive. After two years of collecting data, an n of 100 participants at test and retest had still not been obtained. Although 120 participants had taken the test, only about three quarters that number returned for the retest. However, in light of the Central Limit Theorem, the number of participants is robust enough for the purposes of this study.

Materials

Informed Consent

Prior to taking part in the experiment, each potential participant was asked to sign an informed consent document (Appendix B). This document provided a description of the study, articulated the inclusion criteria, disclosed what the participant was asked to do in the study, and provided information concerning confidentiality and privacy, and contact information. No participant was allowed to participate in the study without first signing the informed consent document.

Sexual Interest Assessment

The test that was used for this study was the Affinity 2.0, a computer program developed to examine sexual interest (Glasgow, 2003). The original version of the Affinity was designed specifically for working with males with learning disabilities. Version 2.0 is licensed for use as a clinical assessment tool with learning disabled adult male offenders and non-learning disabled adult male sex offenders. Glasgow (2003) has also approved the Affinity 2.0 to be used for research and evaluation purposes with adult male non-offenders, juvenile male offenders, and female offenders.

The Affinity 2.0 consists of 10 main parts: 1) The main screen is a screen of identification of the professional user of the instrument. 2) The stimulus management screen which allows the user to determine which available images will be used as practice items and in what order. 3) The clicker screen is a simple activity that serves as an assessment of simple motor skills that may intrude on accurate data reporting. 4) The assessment screen is where basic information about a proposed assessment is entered. 5) The ranking screen provides simple prototype line drawings of males and females of different ages 6) The rating screen is where the participant view a number of images on the computer screen. The fully clothed models depicted in non-sexually explicit poses have been carefully selected to fall within the following groups: adult males, adult females, pubescent males, pubescent females, pre-pubescent males, pre-pubescent females, and male and female children. Each one of these categories corresponds to a prototype line drawing as presented in part 5. The participant is asked to rate each image for sexual attractiveness/unattractiveness. During this procedure, two measures of viewing time are recorded. 7) The results screen displays a table of the assessments

undertaken with the option for viewing all the data gathered on any one of them. 8) The raw data chart screen displays the raw data derived from any individual assessment in the form of a table or a bar chart. 9) The mean ranks screen shows the data (converted to ordinal data) from results using shared axes. 10) The data management screen exports data from any number of assessments for further statistical analysis (Glasgow, 2003).

The participant begins by viewing and ranking several prototype images that are presented in Step 5 of the Affinity. These prototype images are simple line drawings that depict a character from each of the 8 categories. The eight categories are as follows: Adult Male and Female, Juvenile Male and Female, Pre-juvenile Male and Female, and Small Child Male and Female. The participant will begin by ranking the line drawings according to their level of attractiveness. As the participant reaches a point where the figures are no longer attractive to him, he will then begin to rank the remaining figures according to their unattractiveness. The purpose of this prototype ranking procedure is to predict the order of each category when these are ranked either by viewing times or the ratings of attractiveness pertaining to the individual images present in the subsequent rating procedure. Ultimately, this initial ranking procedure is designed to serve as a test for honesty of self-reports when compared to viewing time results (Glasgow, 2003).

The rating procedure consists of showing the participant a total of 56 test images and several practice images. Each of the prototype categories represented in the ranking procedure is made up of seven images. The participant is then instructed to view the picture and then rate the image's sexual attractiveness by using a continuous sliding scale going from "attractive" to "unattractive." As the participant is undertaking this rating procedure, two measures of viewing time are being covertly recorded. The first measure

of viewing time is On Task Latency (OTL) which is the time of first presentation of the image to the time the participant rates the image. The second viewing time recorded is the Post Task Latency (PTL) which is from when the individual rates the image to the time the image is changed. A total viewing time can then be calculated by adding the two viewing times (OTL + PTL). All viewing time measurements are reported in raw score form in milliseconds (Glasgow, 2003).

One of the major benefits of the Affinity 2.0 is that it offers to report all scores in their raw data form rather than having the scores converted into ipsative z-scores as with the AASI (Fischer & Smith, 1999). In order to really determine how heterosexual males responded to visual stimuli, it was essential that we had access to all the raw scores for each individual.

DDSQ

Participants were given a brief questionnaire called the Demographics, Social Desirability, and Sexual Interest Questionnaire (DDSQ) specifically designed for the purposes of this study (see Appendix A). The questionnaire consisted of three sections. The first sections dealt with simple demographics (age, ethnicity, year in school, marital status). The second section was comprised of a social desirability scale named the M-C 2(10) developed by Strahn and Gerbasi (1972) which is a truncated version of the Marlowe-Crowne Social Desirability Scale (M-C 33) (Strahan & Gebrasi, 1972). Strahan and Gerbassi developed two shorter versions of the M-C 33 called the M-C 1(10) and the M-C 2(10); however, pilot testing of the measures showed that the the M-C 2(10) was less offensive and more clearly worded (Mandell, n.d.). The purpose of using a social desirability scale such as this was to compare the social desirability of our sample to a

normative sample. As our sample consists largely of conservative Christians, and as sex is a sensitive subject, the use of this social desirability scale seems very pertinent.

The third section consisted of questions designed to determine the participant's sexual orientation. The sexual orientation inventory included in the DDSQ is an adaptation of the Kinsey Heterosexual-Homosexual Scale (Kinsey, Pomeroy, & Martin, 1998). Since this test is designed to measure the sexual interest of exclusively heterosexual males, those with homosexual leanings were rewarded similarly to the heterosexual experimental sample, but were not included in the sample.

Completion of the DDSQ took place after the administration of the measure of sexual interest in an effort to safeguard the validity of the test. Had the participants completed the DDSQ prior to completing the measure of sexual interest, they might have been sensitized to the fact that those with homosexual tendencies were being screened out. Thus aware, it is possible that they would have completed the questionnaire in a socially desirable manner, as opposed to a truthful manner. Furthermore, although it is beyond the scope of this particular study, the results of all homosexual participants were held for future research. All who participated in the experiment at some level were rewarded with two movie tickets to a local theater.

Procedure

Confidentiality

Since sexual interest is such a sensitive topic, the administration of the Affinity 2.0 and the questionnaire could be considered somewhat intrusive. As such, some students may have been inhibited in their responses unless measures were taken to assure their confidentiality. Brigham Young University is a private religious institution and one

that requires strict adherence to an honor code. This honor code, which all students must sign if they are to attend, prohibits extramarital sexual activity. Failure to adhere to the principles of the Honor Code can result in university sanctions and possible dismissal (Brigham Young University, n.d.). Given that adherence to this honor code is so highly valued, it was the concern of the researchers that students with homosexual feelings might have been reluctant to truthfully respond to the questionnaire and/or the Affinity 2.0. The informed consent document that each potential participant read and signed informed the participants as to the purpose of the study and expectations. Also included was a section that discussed the confidentiality of the identity of the potential participant. This section assured the potential participant that all personal information was to be kept confidential and that no names would be used in the study nor reported to the Honor Code Office of Brigham Young University. This informed consent document was designed to protect the individual, lessen his inhibitions about being a participant, and increase the chances for honest responses on both the questionnaire and the Affinity 2.0. To further safeguard against breach of confidentiality, participant's names were not used after the initial data had been gathered. To ensure the confidentiality of the participants, each was assigned a number that served as their identification as the data were analyzed. The master list of participants' names and numbers was kept in a lockbox to further ensure confidentiality.

Setting

Another area that might have affected participants' viewing time was the actual setting in which they took the Affinity 2.0. Martin (1964) showed that individuals asked to rate sexually explicit photos spent much less time looking at those photos in the

presence of other raters than if they were alone. Brown, Amoroso, Ware, Pruesse, and Pilkey (1973) looked at the factors affecting viewing time of pornography. They hoped to find a change in viewing time as the explicitness of the sexual stimuli increased. Male participants were asked to rate a total of 15 slides of varying degrees of sexual explicitness. Participants were informed that they could look at the images as long as they liked. Participants were either alone or observed by three graduate students that reported being interested in the study. Results showed that the participants spent significantly less time looking at the sexually explicit slides in the presence of others than when they were alone.

The aforementioned studies differed from this one in that participants were asked to rate sexually explicit photos as opposed to clothed models in non-explicit poses. However, participants in this study did know that the test they were taking was a measure of their sexual interest. If unaccounted for, this knowledge might have led the participant to reduce his viewing time in the presence of other individuals. Thus, we decided to assess each participant individually rather than in groups.

Administration of Instruments

After having read and signed the informed consent documentation, the participants were led to a private room that was equipped with a single computer on which the Affinity 2.0 program was installed. The researcher instructed the participant pertaining how to start the program, and helped the participant complete the primary prototype ranking procedure and supervised the completion of the rating of the sample images as per the instructions given by the Affinity 2.0 manual (Glasgow, 2003). The researcher then left the room and allowed the participant to view and rate the test images.

The researcher instructed the participant that he or she would wait outside to make sure that no one would disturb him. This setting assured the participant's privacy, which should have reduced inhibitions and promoted honest reactions as well as more accurate viewing times. Once completed, the participant would exit the room and the researcher could answer any additional questions and schedule a second visit for two to four weeks from the time of the initial assessment. The researcher then entered the room, made sure that the data had been recorded, and prepared the room for the next participant. Once the participant returned to be re-tested after the two to four week interim, he followed the same procedure, with the added necessity of now filling out the DDSQ questionnaire after completion of the Affinity 2.0. Upon completing the Affinity 2.0 for a second time and filling out the DDSQ, the participant was given two single admittance movie tickets to a local theater as a reward for his participation. Every participant involved in this study followed this exact procedure.

Data Analysis

Descriptive Statistics

The purpose of this study was two-fold; to ascertain whether or not there is a characteristic curve to the eight categories of the Affinity 2.0, as well as to test the temporal stability of this instrument. The data in this study is most easily viewed as existing in a three-dimensional cube. The Affinity 2.0 consists of eight categories, each comprised of seven slides. The test was administered to 88 participants on two separate occasions in an effort to examine the temporal stability of the instrument. Although the scores for each slide are given in raw data form, dealing with such a large amount of raw data (11,200 cells) would be cumbersome and unnecessarily time-consuming. Thus, we

determined to use one score for each category that represents the raw data included in the seven slides. This method limited the number of data to 1,600 cells, thereby making analysis of the data much more efficient. At the outset it was not immediately evident exactly which score should be used to best accomplish these purposes; therefore we tested a variety of statistical methods for arriving at the most meaningful scores as well as the best measure of temporal stability.

Sums. A simple summation of the viewing times for the seven slides in each category was used as an effective representation for that category. Summing the scores was extremely simple and proved to be much more efficient than listing each of the seven viewing time scores independently. As summing of VT scores was chosen as a representation of each of the eight categories, the average sums and standard deviations were used to represent the distribution of scores.

Means. Averaging the VT scores from each of the eight categories also served as an effective representation of a given category. Obtaining the mean simply required adding each of the seven VT scores, and then dividing by seven. The distribution of each of the eight categories was represented by the average means and standard deviations.

Medians. Medians functioned as another effective representation of the seven VT scores that avoided some of the pitfalls of using mean scores. It is possible that some participants may have viewed a given slide for an unusually long time for a reason other than that they are sexually interested in the person depicted. For example, an adult male may have viewed a slide of a young male child for an extended period of time because he looked like his own child of the same age. For this reason and many others, it is possible that some participants may have looked at a given slide for an extended period of time for

reasons unrelated to the construct being measured. Since they are greatly affected by skewed distributions, the use of means in such cases would have painted an inaccurate picture of the person viewing the slides. Medians avoided this problem by calculating the fiftieth percentile rank rather than the simple average. In an effort not to skew the test results, the use of medians as representations of each of the eight categories was explored. When medians were used as the representation of the eight categories, the shape of the distributions was represented using the medians and interquartile ranges.

Ipsative Measures

Cattell (1944) has stated that there are three types of psychological measurement of behavior: raw, normative, and ipsative. Raw data are neither dependent on any other scores of the individual measured nor upon the scores of any other individuals . In this study, had we used all seven VT scores, on all eight categories, for all 100 participants; we would have used their raw data. Normative units of measure are where the score of the individual is dependent on the scores of others in the population. The three aforementioned methods (sums, means, medians) for arriving at a representation of the seven VT scores are all examples of normative measures. Finally, ipsative units are scores for a participant that are dependent upon his scores on other variables (Clemans, 1956). The final three methods we employed as representations of the VT raw scores are all examples of ipsative measures.

As defined by Horst (1963) ipsatives are, “any score matrix, which has the property that the sum of the scores over the attributes for each of the entities is a constant.” It is possible to ipsatize any set of scores by adding a suitable constant to the measure of each attribute for a specific entity such that the new scores will sum to the

same constant for all entities (Clemans, 1956). Before ipsatizing scores it is crucial that the scores first be standardized before the ipsatization process takes place. Failure to standardize the scores will result in "bastard ipsative measures" that are effectively devoid of meaning (Clemans, 1956). The major advantage of ipsative scores is that they allow for intraindividual comparisons to be made. Once ipsative scores have been calculated, strengths of the attributes for an individual can be compared to each other by ranking his ipsative scores, providing the means and variances of the attributes were calculated prior to ipsatization.

While they are not as widely used as more traditional measurement methods, we initially felt that ipsative measurements might show promise when calculating raw scores from the Affinity 2.0. It was hypothesized that intraindividual comparisons made possible by ipsative measures might have been of special importance when working with the highly sensitive topic of sexual interest. For example, it may appear that a given participant looks for an unusually long time at pictures of small children of the same sex. While this may initially appear troublesome, it is possible that ipsatization of the participant's score may show that their VT scores are not unusual given their total VT pattern. As we were dealing with such a sensitive issue, the researchers felt it was imperative to have statistical methods that carefully analyzed what was truly being measured.

Ranks. The first ipsatization of the raw scores was to convert them into ordinal ranks. Each of the eight categories was given an ordinal rank based on the mean VT for the seven slides within that category. The distribution was then represented by the standard deviations of the ordinal ranks. Working with ordinal scales in this manner has

been advised against by Barrett (2002), but is used in many other statistical analyses such as the Mann-Whitney U test (Bruning & Kintz, 1997). At the outset of this study, it was unclear whether the use of ordinal ranks would be effective at describing the eight categories. However, in the spirit of exploration, the researchers evaluated the effectiveness of this procedure in our search for the most meaningful category descriptor. A description of the effectiveness of this approach is found in the Results section.

Ipsative Weighted Ranks. A more sensitive ranking than the above-mentioned ordinal ranks, is ipsative weighted ranks. Ipsative weighted ranks use larger numbers than ordinal ranks, making it possible to notice slighter variations. Additionally, ipsative weighted ranks are already calculated by the Affinity 2.0, making them useful, easily calculated and readily available. The Affinity 2.0 performs this calculation by first assigning a value to each of the 56 slides according to how long it was viewed. For example, the slide looked at longest by the participant will be given a rank of one; conversely, the slide looked at for the shortest period will be assigned the number 56. The sum of all of these ranks equals 1,596. Next, the ranks of each of the seven slides in all eight categories is summed and divided by seven.

Temporal Stability

The second purpose of this study was to establish whether or not the Affinity 2.0 is a temporally stable measure of sexual interest. It was hypothesized that sexual interest is a relatively stable characteristic, so it seemed reasonable to assume that a valid measure of sexual interest would yield similar results on test-retest trials. Temporal stability was measured using Pearson Product Moment Correlation Coeffecients for interval data and Spearman's Rank Correlation Coefficients for ordinal data. Chi- square

also showed promise as a measure of temporal stability. A Chi- square analysis was used in this study, with mean observations from time one serving as the expected scores and scores from time two serving as the observed scores. Scores that were not significantly different from one another, were seen as evidence that the instrument was temporally stable.

Summary

Since the Affinity 2.0 is a relatively new instrument, we felt it advisable to try a variety of analyses of the raw data provided by the instrument. Because this research was basic research and was without precedent, we tried each of the aforementioned six methods for analyzing the raw data mentioned and reported on the effectiveness of each method. It is our hope that doing so has provided valuable direction for future researchers using the Affinity 2.0.

Chapter 4

Results

Participant Demographics

Participants in the study were to take the Affinity 2.0 and return after roughly two weeks for a second administration of the instrument. After completing the Affinity 2.0 assessment process for a second time, participants were given a questionnaire (Demographic Data Survey Questionnaire or DDSQ) with questions regarding age, ethnicity, year in school, marital status, reference source, sexual orientation, as well as a measure of social desirability. These data are useful in painting a picture of the participants involved in this study, and highlight some of the strengths and limitations of the study itself.

One hundred and twenty male participants completed the first administration of the Affinity 2.0. Of these 120, 88 returned, retested, and filled out the Demographic Data Survey Questionnaire, thereby fulfilling the inclusion requirements of this study. The mean age of participants in this study was 22.98. Participants in the study ranged in age from 18 to 45 years old. All of the participants in this study were college students, with each year being represented. Twenty-five of the participants were freshmen (28.4%), 13 were sophomores (14.8%), 19 were juniors (21.6%), 26 were in their senior year (29.5%), and five were graduate students (5.7%).

Twenty-three (26.1%) of the participants in the study were married and the remaining 65 students (73.9%) listed their marital status as “single.” Although “divorced” and “widowed” were both options on the DDSQ, none of the participants endorsed these items. Recruitment for this study occurred largely in undergraduate psychology classes, a

fact that is reflected in the makeup of the participants. Of the 88 participants who completed all inclusion requirements, 66 (75%) became aware of the study in a psychology class. The remaining 22 students (25%) were recruited largely through "word of mouth" efforts by friends who had previously participated in the study.

Consistent with the composition of the Brigham Young University student body, a majority of the participants in the study listed their ethnicity as "White" or "Caucasian." Eighty of the 88 participants in the study belonged to this category, accounting for (90.9%) of the total participants. Of the remaining eight participants, two (2.3%) listed their ethnicity as "Asian," one (1.1%) as "Northern European," one as "Mexican/Samoan", one as "Pacific Islander," one as "Hispanic," one as "Western Indian," and one as "Native American."

Finally, the Marlowe-Crowne Social Desirability Scale-10 was included in the DDSQ as a means of assessing the participants' propensity to answer personal questions in socially desirable ways. The average score on the Marlowe-Crowne Desirability Scale-10 is 4, indicating 4 socially desirable answers out of the 10 possible questions. Participants in this study averaged 4.079 on the social desirability scale. While this score is slightly above average, it is still within the normal range and presents no significant threat to the validity of the study.

As this study is seeking to provide a basal measurement for the sexual interest of heterosexual males, participants who defined themselves as anything other than "exclusively heterosexual with no homosexual interest" (1 on the Kinsey Scale) were not included in the final data analysis. Although all of those included in the study ultimately endorsed the "exclusively heterosexual" box on the Kinsey scale, 3 of the participants

initially indicated having some homosexual experience (2 or 3 on the Kinsey Scale in this case) before marking out their answer and foreclosing on the "exclusively heterosexual" option.

Sums

It was initially proposed that three separate analyses be run based on the On Task Latency (OTL), Post Task Latency (PTL), and Total Task Latency (TTL) scores of the participants. The initial hypothesis was that each of three measures would provide different information that would give a more holistic picture of the participants' sexual interest. Further analysis of these constructs did not support the original hypothesis. Having found PTL and TTL to be redundant and uninformative in completing the aims of this study, they were eliminated from our analyses.

Kara Harmon (2006), in conducting the female analogue to this study, found that the average range of PTL across images/categories was .76-.91 seconds- an average of only .15 seconds. Brian Sneed (2006) also found that the average PTL scores in his sample of men was between .68 and .94 seconds. Post Task Latency is a measure of the amount of time subjects spent gazing at a picture following completion of the rating task. Our original hypothesis was that subjects might linger on slides they found sexually attractive, even after having completed the assigned rating. Had this been the case, PTL would have provided another unique measure of sexual attraction that would have been germane to our analysis. However, taking Kara Harmon's findings into consideration, it would appear that participants did not routinely linger after rating the slide, and that PTL accounted for a very small amount of time and was consistent across slide categories.

Total Task Latency was a construct that is not currently included in Affinity 2.0 software. Total Task Latency is derived by summing On Task Latency and Post Task Latency scores (OTL+PTL). Our initial hypothesis was that Total Task Latency scores would provide valuable information that might be more indicative of true sexual interest than either OTL or PTL alone. Unfortunately, given the consistently small amount of time spent on Post Task Latency, adding these scores to On Task Latency scores did not significantly alter or improve them.

Measures of Post Task Latency were found to be similar across slide categories and were consistently of no use in providing a meaningful difference from On Task Latency scores. Inasmuch as these scores do not aid in establishing the normal heterosexual male response to the Affinity 2.0, they have been omitted from subsequent analyses. Included in this omission, is a comparison of the sums of OTL, PTL, and TTL scores. On Task Latency scores served as the basis of the remaining analyses.

Means

Means based on OTL scores were computed at both test and retest. These means can be found in Table 1 and graphical comparisons of test and retest scores can be found in Figure 1. A comparison of the test and retest results shows two clear patterns: a decrease in overall viewing time from time one to time two and a clearly discernable viewing time preference for adolescent and adult women. As can be seen in Table 1, each of the eight categories had an overall shorter viewing time at retest. This phenomenon can be readily explained in terms of the participants' familiarity with the instrument and the task required of them. Participants were initially asked to perform a task unfamiliar to

them, and their viewing times decreased as they gained greater familiarity with the slides, the method of administration, and the rating task.

Table 1

Mean (OTL Total)

Category	Test			Retest		
	Mean	SD	Range	Mean	SD	Range
ADF	3.68	1.43	8.29	3.12	1.20	5.93
JUF	3.47	1.51	9.27	2.82	1.10	5.43
PJF	2.03	1.05	6.63	1.69	0.78	4.23
SCF	2.01	1.13	6.99	1.58	0.55	2.65
ADM	2.23	1.32	9.04	1.75	0.84	4.56
JUM	2.10	1.31	7.07	1.61	0.68	4.14
PJM	1.86	0.97	6.18	1.57	0.64	3.50
SCM	1.71	0.92	4.59	1.46	0.50	3.09

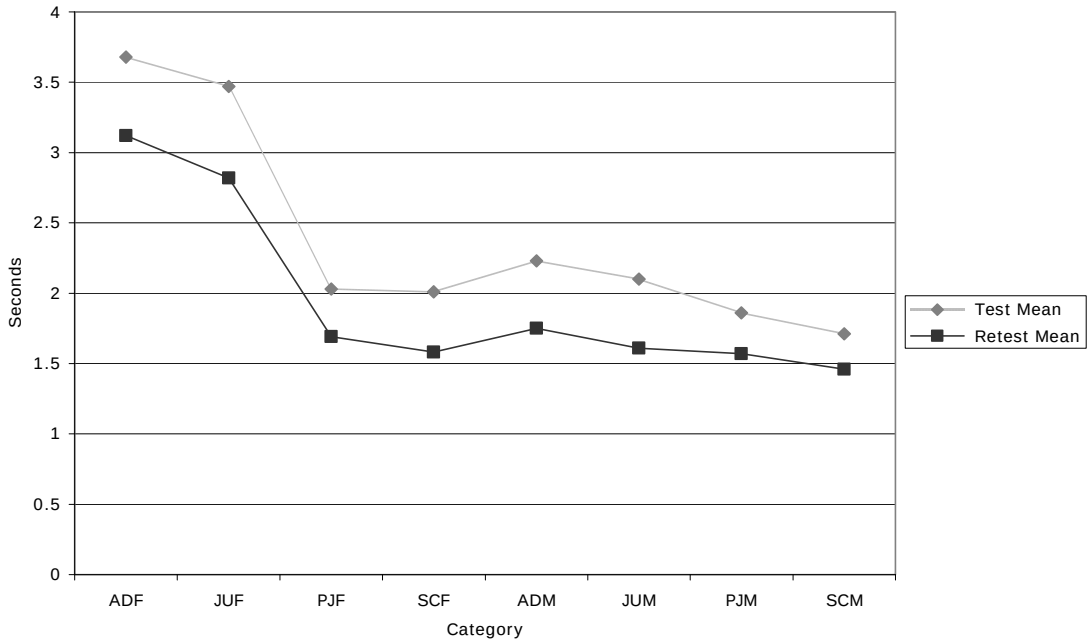

Figure 1 . Mean (OTL Total)

At both test and retest, images of adult females (ADF) and adolescent females (JUF) were the clearly preferred visual stimuli. Slides of adult males (ADM) were the third longest attended to visual stimulus at both administrations of the instrument.

However, it should be noted that there is a precipitous drop in viewing times between the second and third most preferred slide categories. On average, participants spent 3.4 seconds viewing ADF slides, and 3.14 seconds viewing JUF slides. Although ADM slides were viewed third longest, the average seconds spent viewing is 1.99, a significant drop from the previous two categories. At test, the remaining categories were as follows, listed in descending order of viewing time preference: adolescent males (JUM), pre-juvenile females (PJF), small child female (SCF), pre-juvenile male (PJM), and small child male (SCM). At retest, the ordering of the preferred viewing time categories is slightly different, with PJF being fourth most viewed followed by JUM, SCF, PJM, and SCM respectively.

Although the ordering of the viewing time categories is slightly dissimilar between test and retest, a clear viewing time preference for adult and adolescent females is shown in both administrations of the instrument. This consistent viewing time preference for depictions of nubile females, paired with the dramatic decline in viewing time scores for slides of males and children seems to suggest that this curve is representative of a normal heterosexual male response to the Affinity 2.0. Further evidence of this is that the largest standard deviations in viewing time at both test and retest can be found in the ADF and JUF categories. Smaller standard deviations among the male and younger children categories suggest a lack of viewing behavior that further supports our assertion that this sample represents a normal heterosexual male response to this instrument.

Pearson Product Moment Correlation as well as Spearman's Rho coefficients were calculated to assess the consistency of viewing time behavior from test to retest.

Without exception, the Pearson Product Moment Correlations were more robust and will be reported in lieu of the Spearman's Rho data. The results of this analysis can be seen in Table 2 below. All coefficients were statistically significant at the $p < .01$ level, although the strength of the correlations differed between categories. Some categories demonstrated only mild to moderate practical significance (ADF, JUF, PJM, SCF), while others showed moderate to strong practical significance (ADM, JUM, PJF, SCM). These scores were significant across categories, but the strength of this correlation was largely category dependent.

Table 2

Pearson Product Moment Correlation of Affinity 2.0 Mean– Test to Retest

Category	ADF	JUF	PJF	SCF	ADM	JUM	PJF	SCF
r	.434**	.431**	.743**	.391**	.668**	.624**	.743**	.391**

** Correlation is significant at the .01 level (2-tailed)

Medians

In addition to means, medians were calculated as a second measure of central tendency of the viewing time scores. In instances of skewed data, the median serves as a more reliable measure of central tendency and may prove more useful than the mean. The presence of skew can be detected by comparing mean and median scores. If, on average, medians are higher than means, negative skew is present. Conversely, greater mean scores suggest the existence of positive skew. A comparison of the mean and median scores of this study revealed that mean scores were greater than median scores, indicating positive overall skew. A closer examination of the Affinity 2.0 reveals that the very nature of the instrument may lend itself to a likelihood of skew. As a result of the low average viewing time for all slides, as well as the relatively small number of slides, skew is very likely. In light of the likelihood of skew, median scores are likely the best measure

of central tendency for this and future administrations of the Affinity 2.0. Table 3 illustrates the average median OTL, as well as the standard deviation and range for each category. Figure 2 represents the test and retest median viewing times for each of the eight categories.

As was the case with the means, median viewing time scores dropped from the first to second administration of the Affinity 2.0. Another similarity is the clear preference for ADF and JUF images, and the marked drop when arriving to viewing time scores of men, boys, and younger children. Participants spent an average of 3.62 and 3.15 seconds viewing slides of adult and juvenile women respectively. Slides of adult men, the third longest viewing time category, averaged a mere 2.01 seconds per slide. While differences of 1.61 and 1.14 seconds may not appear substantial, they actually constitute a very great difference given the relatively short amount of time spent viewing each slide.

Median scores provide a slightly different ordering of viewing time categories than do the means. At test, both mean and median scores provide similar results, with ADF, JUF, and ADM as the three longest viewed categories (in the order presented). Utilizing means, JUM is the fourth longest viewed category, while it ranks sixth using the median. Median scores yield PJF, SCF, JUM, PJM, and SCM respectively, as the fourth through eighth categories as measured by viewing time. Thus, it would appear that the measure of central tendency utilized may have an important impact on the relative ordering of the categories by viewing time. Median measures of retest conform more closely to the mean measures than did the first administration. As with the mean, the five longest viewed categories (in descending order) are ADF, JUF, ADM, PJF, and JUM. There is a slight difference in the final three categories with median rankings yielding

Table 3

Medians

Category	Test			Retest		
	Mean	SD	Range	Mean	SD	Range
ADF	3.96	1.71	10.20	3.27	1.41	7.15
JUF	2.07	1.15	7.08	1.77	0.81	4.24
PJF	2.18	1.16	6.30	1.83	0.81	4.87
SCF	1.92	1.13	8.48	1.65	0.66	4.14
ADM	3.51	1.54	9.23	2.78	1.01	4.80
JUM	2.02	1.17	8.33	1.60	0.49	2.80
PJM	2.02	1.21	9.35	1.70	0.66	3.91
SCM	1.69	0.70	4.47	1.55	0.49	2.79

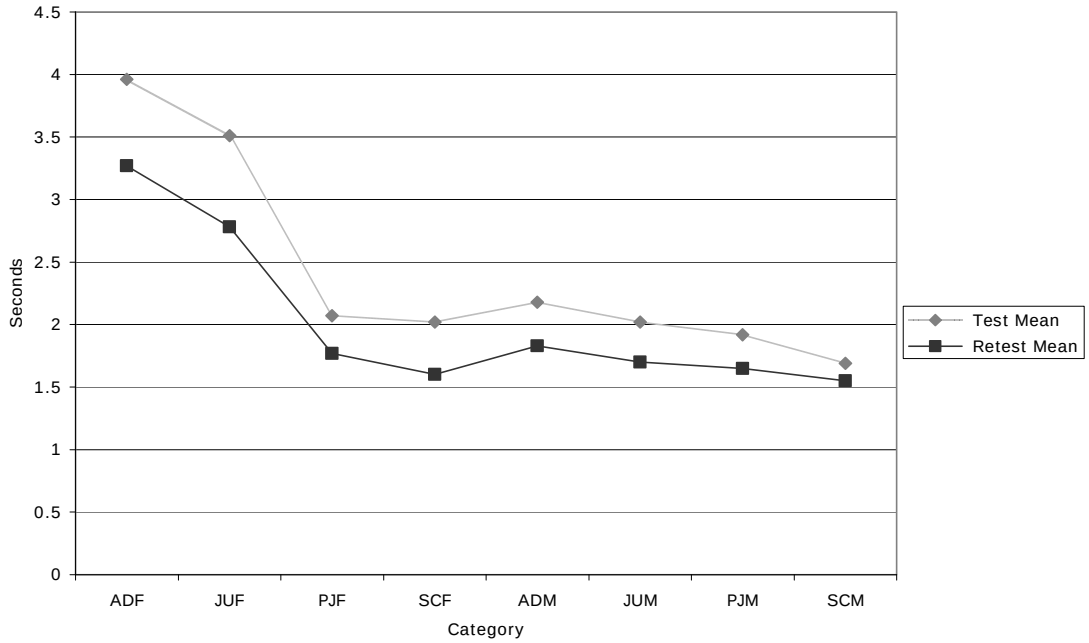

Figure 2. Mean Medians

PJM, SCF, and SCM as the three least viewed; as opposed to SCF, PJM, and SCM for the means.

As with the mean data, both Pearson Product Moment Correlations and Spearman's rho were used to analyze the median scores. In an effort to maintain consistency, and also given the generally stronger findings of the Pearson Product Moment Correlation, these data will be reported herewith in lieu of Spearman's rho. The aforementioned results can be found in Table 4 below. Without exception, all correlations were significant at the $p < .01$ level. As with our analysis of the mean scores, the correlations of the pre-post medians exhibited a range of strengths from mildly correlated, to strongly correlated, dependent upon category. Three categories (ADF, SCF, PJM) showed correlations that were in the mild to moderate level. This suggests quite a bit of variance in the consistency of the participants' responses from time on to time two, relative to the responses of other participants. The remaining five categories (JUF, PJF, ADM, JUM, SCM) showed stronger correlations, suggesting a good deal of consistency in viewing time behavior between administrations of the test, as measured in relation to the responses of other respondents.

Table 4

Pearson Product Moment Correlation of Affinity 2.0 Medians – Test to Retest

Category	ADF	JUF	PJF	SCF	ADM	JUM	PJM	SCM
r	.324	.636	.691	. 298	.719	.800	.427	.597

Note. All correlations significant at the 0.01 level (2-tailed).

Ipsative Scores

Category Ranks

An analysis of category ranks was one of the ipsative procedures originally agreed upon in the prospectus. The first step in obtaining the category rank is to sum the time spent by each participant gazing at each of the seven images within the eight categories. Having thusly obtained the sums for each of the categories, the categories were then ranked 1 through 8. A score of "1" represented the category in which the most time was spent viewing the slides, with "8" representing the category where the least amount of time was expended. Finally, a Pearson Product Moment Correlation was conducted, comparing the each individual's scores from time test and retest.

Table 5

Pearson Product Moment Correlation of Affinity 2.0 Category Ranks – Test to Retest

Category	ADF	JUF	PJF	SCF	ADM	JUM	PJM	SCM
r	.263*	.034	.369**	.077	.296**	.133	.008	-.012

** Correlation is significant at the .01 level (2-tailed)
* Correlation is significant at the .05 level (2-tailed)

As evidenced by Table 5, only three of the eight categories were found to be statistically significant. Two categories (ADM, PJF) were significant at the .01 level and one (ADF) was found to be significant at the .05 level. Despite the statistical significance of these three categories, each of the correlations is fairly weak. The remaining five categories were even more weakly correlated with one category (SCM) actually having a negative correlation. Given these findings, it would appear that the transformation of raw OTL viewing times into category ranks results in too much distortion to make this a useful descriptor of our data. One possible explanation for this distortion is that the

assignment of scores 1-8 to the categories is too broad, and fails to capture some of the nuanced differences in scores.

Affinity 2.0 Ipsative Weighted Ranks

The next analysis was the Affinity 2.0 ipsative weighted ranks. The results, including mean, standard deviation and range for both test and retest can be found below in Table 6.

Table 6

Mean Ranks (Weighted Ranks)

Category	Test			Retest		
	Mean	SD	Range	Mean	SD	Range
ADF	43.50	7.20	40.15	44.16	6.97	29.28
JUF	40.43	6.23	24.79	40.18	6.58	34.14
PJF	25.00	5.85	28.85	25.22	6.81	31.00
SCF	24.14	5.72	27.07	23.51	5.22	28.71
ADM	27.51	6.37	31.79	26.40	7.75	37.71
JUM	25.18	5.36	24.86	24.70	4.60	22.57
PJM	23.11	5.12	27.79	22.92	4.88	22.50
SCM	18.93	5.19	33.28	20.58	5.68	27.76

As can be seen in Table 6, there are notable similarities between the time one and time two ranks. At time one, ADF was ranked highest, followed by JUF, ADM, JUM, PJF, SCF, PJM, and SCM respectively. Time two ranks were only slightly different, from highest to lowest ranked: ADF, JUF, ADM, PJF, JUM, SCF, PJM, and SCM. At test, JUM was ranked fourth highest with PJF immediately behind; whereas retest yielded PJF as the fourth highest ranked category, with JUM immediately following. In both cases, the differences were minute, amounting to less than one second in each instance. What is more significant is that, as measured by the Affinity 2.0 weighted ranks, our subjects showed a clear and significant preference for adult and juvenile females. In addition to

the similarities between the means at time one and time two, standard deviations and range scores were also very similar.

Male Mean Ipsative Weighted Rank Response Curve to Affinity 2.0

Of significant interest was the similarity between the ipsative weighted rank response curve and the mean OTL raw score response curve. Unlike the category ranks, this measure maintains the integrity of the curve as well as the ordering sequence of the preferred categories. This being the case, the ipsative weighted rank response measure seems to be useful as a means of describing the participants' response to the Affinity 2.0.

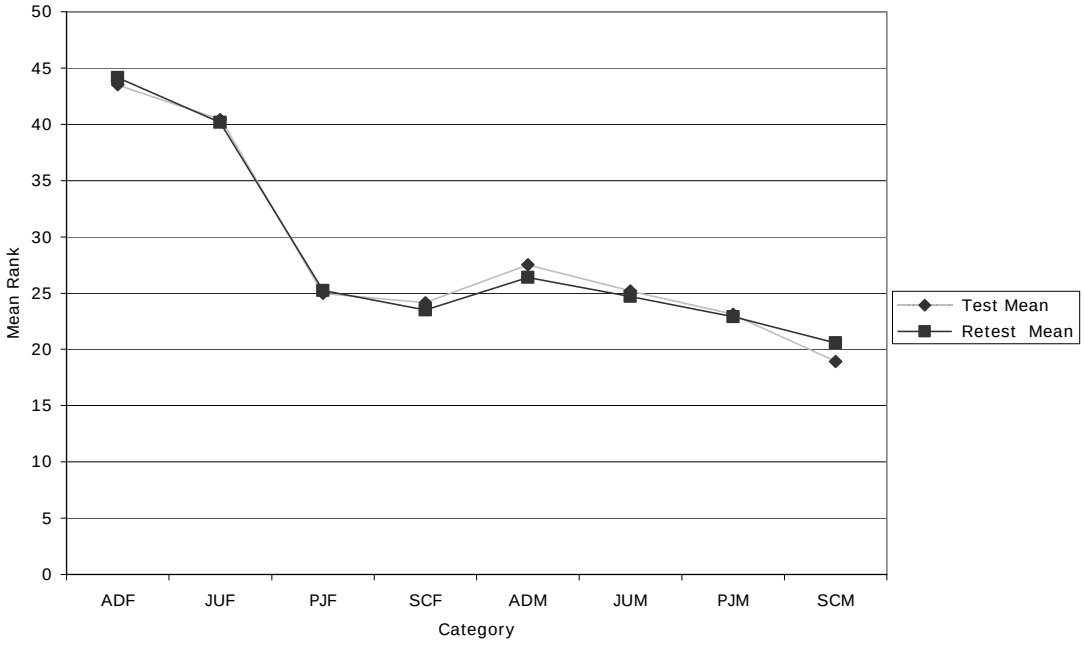

Figure 3. Mean Ranks (Weighted Ranks)

The above figure (Figure 3) juxtaposes the mean ranks responses from time one and time two. As mentioned above, the order of preferred categories was slightly different from time one to time two. However, the differences between time one and time two were slight. The weighted ranks being discussed here are ipsative measures of intra-

individual response, which differentiates them from other measures included in this study (e.g. – OTL means).

Test/ Retest

Pearson Product Moment Correlations were calculated for the ipsatized weighted mean ranks and are shown below in Table 7. As seen below, the correlations for ADF, PJF, ADM, and JUM were all statistically significant at the 0.01 level. Of the significant categories, only ADF and ADM were approaching moderate strength, with PJF and JUM being only mildly strong. The remaining categories (JUF, SCF, PJM, and SCM) did not yield statistically significant correlations.

Table 7

Pearson Product Moment Correlation of Affinity 2.0 Ipsative Weighted Ranks – Test to Retest

Category	ADF	JUF	PJF	SCF	ADM	JUM	PJM	SCM
r	.487**	.196	.392**	. 123	.535**	.440**	.005	.201

** Correlation is significant at the .01 level (2-tailed)

These results suggest that the individual image rankings within categories tended to vary more than did the rankings of categories at large. Overall, the images within the adult categories (ADF, ADM) were the most consistently ranked, with the most variability occurring in the pre-juvenile and small child categories. Of noted exception to this is the PJF category, which showed a mildly strong correlation of .392 that was significant at the 0.05 level.

Table 8

OTL Conversion Affinity 2.0 Chi-Square Estimate of Temporal Stability – Test to Retest

	ADF	JUF	PJF	SCF	ADM	JUM	PJM	SCM	Total
Observed	43.50	40.43	25.00	24.14	27.51	25.18	23.11	18.93	227.80
Expected	44.16	40.18	25.22	23.51	26.40	24.70	22.92	20.58	227.67
O-E	-0.66	0.26	-0.23	0.63	1.11	0.48	0.19	-1.65	
O-E SQ	0.44	0.07	0.05	0.39	1.23	0.23	0.03	2.71	
O-E SQ/E	0.01	0.00	0.00	0.02	0.05	0.01	0.00	0.13	0.22
Residual	-0.10	0.04	-0.05	0.13	0.22	0.10	0.04	-0.36	

Chi-Square for Temporal Stability

The final analysis proposed as part of the prospectus was a Chi-Square that would serve as an estimate of the temporal stability of the Affinity 2.0. To compute the Chi Square, average test (time one) results were used as the expected scores, with average retest (time two) scores used as the observed values. Chi-square values under 14 (df =7) with no standardized residuals exceeding $\pm$ 1.96 would indicate that the Affinity 2.0 is temporally stable.

Our original hypothesis was that sexual interest would be stable across both administrations of the Affinity 2.0, and that none of the residuals would surpass the significant $\pm$1.96 level. To test this hypothesis, the Chi-square goodness of fit test was run comparing weighted ranks from the two administrations of the instrument. As seen in Table 8, the Chi-square value of 0.219 (df=7) was obtained which was not statistically significant. None of the standardized residuals surpassed the significant $\pm$1.96 level.

Chapter 5

Discussion

One decided advantage of using the Affinity 2.0 as opposed to extant viewing-time measures, was that the Affinity 2.0 afforded us direct access to participant data. This data allowed computation of the various analyses that eventually allowed for inter-individual comparisons. Having access to this data also allowed additional analyses as needed and obtain valuable information regarding the participant responses to the Affinity 2.0.

Summary of Results

The purpose of this study was to address two specific questions. First, is there a characteristic pattern to normal heterosexual males viewing time responses to the Affinity 2.0? Second, is the Affinity 2.0 temporally stable? The Affinity 2.0, an instrument that surreptitiously measures viewing time as an indicator of sexual interest was used to address the research questions. Our initial hypothesis was that a characteristic pattern to normal heterosexual sexual interest would emerge that is temporally stable as measured by the Affinity 2.0. Our results indicate that a characteristic curve does exist and that it is reasonably temporally stable.

At test and retest administrations of the Affinity 2.0, slides of adult females (ADF) and adolescent females (JUF) were the clearly preferred visual stimuli of our sample. The participants' consistent preference for slides of nubile women suggests the presence of a characteristic pattern to normal heterosexual males viewing time response to the Affinity 2.0. At both test and retest, ADF slides were viewed longest, with an average viewing time of 3.89 seconds per slide within the ADF grouping. Mean results

for the JUF grouping shows an average viewing time of 3.6 seconds per slide. The third most viewed grouping was adult males, with an average viewing time of a mere 2.28 seconds. This significant drop indicates the sample's clear preference for sexually mature women and is indicative of a characteristic curve. Median scores were also computed, and may provide a more accurate measure of central tendency, given the overall positive skew of the results. Median scores buttress the findings of the mean scores, as ADF and JUF slides respectively were most preferred. Participants viewed ADF slides at a median rate of 3.62 seconds, with JUF slides viewed at the median rate of 3.15 seconds per slide. Consistent with the mean results, our analysis of the median scores showed adult males (ADM) to be the third longest viewed grouping of slides, albeit at a much- decreased rate. Median scores for ADM slides were 2.01 seconds per slide, a significant decline from the preferred ADF and JUF groups. As evidenced by the results of our mean and median analyses, our sample showed a consistent preference for viewing slides of sexually mature adults and adolescent females. Another interesting characteristic of the viewing time curve is that ADM slides were the third most viewed slides as measured by both means and medians. While these slides were viewed for significantly less time than those of the nubile females, and may not indicate sexual interest in males, their presence as the third most viewed category of slides does warrant further attention.

In her analogous study of normal heterosexual female responses to the Affinity 2.0, Kara Harmon (2006) found that females consistently viewed slides of adult males (ADM) longest followed by juvenile males (JUM) and then adult females (ADF). Harmon hypothesizes that this viewing time behavior is indicative of social comparison and hypothesizes that her female participants are examining these ADF slides as a means

for comparing themselves. While this behavior may be more prevalent among women, this social comparison phenomenon may also exist among heterosexual men. Although concerns of attractiveness are stereotypically viewed as a female concern, these findings may indicate the presence of comparative behavior among men as well. Men may be socialized to compete and may engage in comparative behaviors as a means of establishing an internalized hierarchy based on attractiveness. This hypothesis warrants further exploration, as does the general idea of male views on sexual attractiveness.

The second question posed by this study deals with the temporal (test-retest) stability of the Affinity 2.0. Our original hypothesis is that sexual interest, as measured by viewing time response to the Affinity 2.0, would be a stable construct. Two analyses were run to address the question of temporal stability. The first, Pearson Product Moment Correlations (PPMC) showed mild to moderate strength. This means that, as measured by the responses of other respondents within a given category, the individual categories had anywhere from relatively mild to moderately strong levels of temporal stability. Thus, it would appear that some categories may better address the issue of test-retest consistency than others. It is worth mentioning that all of the categories were statistically significant, as measured by the PPMC.

A Chi-square goodness of fit was proposed as the second measure of temporal stability for this study. Time one administration scores were used as the “expected” scores, with time two administrations serving as the “observed” scores. No significant deviation from time one to time two was obtained and none of the residuals were significant at the +-1.96 level, suggesting that the instrument is temporally stable for the purposes of this study. The results of this chi-square analysis suggest that, on average, the

results of the respondents as a whole are temporally stable. These findings suggest that the Affinity 2.0 may provide a temporally stable, non-invasive means of approximating sexual interest.

Theoretical Implications

Further support for the hypothesis that viewing time and sexual interest are closely linked would come by correlating Affinity 2.0 measures of participant sexual interest with their scores on a penile plethysmograph (PPG). Whereas penile tumescence is clearly related to sexual interest, the relationship between vision and sexual interest is potentially confounded by the variety of uses of vision. Vision is paired with a variety of survival functions (e.g. – ambulation), some of which are loosely related or unrelated to sexual interest. Therefore, it is possible that confounding factors may exist within viewing time measures of sexual interest as a result of the myriad functions of vision. However, if a strong correlation was shown to exist between PPG measures of sexual interest and viewing-time measures of sexual interest, it would lend support to the thought that viewing-time has excellent potential for measuring sexual interest.

While our results suggest that increased viewing time is predictive of greater sexual interest, further work needs to be done to test this hypothesis. For instance, our analysis included only males whose self-reported sexuality was “exclusively heterosexual” (1 on the Kinsey scale). To further corroborate our hypothesis that increased viewing time is indicative of increased sexual interest, similar analyses would need to be undertaken with individuals whose self-report was other than “exclusively heterosexual” on the seven point Kinsey scale (2-7). A logical next step would be to study normal, exclusively homosexual men (7 on the Kinsey scale) to see if their results

mirror those of our sample. If the current hypothesis is to be borne out, self-reports of sexuality will closely correspond to viewing time indicators of sexual preference, as was the case with our study.

Of final theoretical interest is an examination of the specific workings of increased viewing time as it relates to measure sexual interest. Specifically, what processes comprise the added time spent viewing the slides found more sexually attractive? In some instances it may be simple appetitive gazing, in other words gazing with some level of increased sexual arousal at an image found to be generally sexually appealing. Another hypothesis, the social cognition model, is that the image may be broken down according to categories of sexual interest specific to the participant. Checking for areas of specific interest (e.g. – hair color) may account for the increased time spent viewing the slide, whereas slides obviously lacking these categories can be more quickly bypassed. The results of this study, which found PTL to be too small to work with, seems to argue against simple appetitive gazing and for social cognition.

Clinical Implications

Perhaps the greatest clinical application of the current data is that they allow for norm-referenced decisions to be made. Historically, researchers have sought to make statements of normality, abnormality and deviance from ipsative scores. In reality, this logic is ill-founded as there existed no normal curve from which to make such comparisons. The current study seeks to provide a temporally-stable, normal curve from which statements of normality and abnormality can be more accurately drawn. The Affinity 2.0 has a number of clinical applications including screening, diagnostic, and prognostic functions. Intuitively speaking, screening cannot occur without the existence

of a normal curve that can be used as a basis for comparison. The current study provides a temporally stable, normal curve for heterosexual males that can be used as the comparative base necessary for screening purposes. Traditional ipsative measures have sought to describe a "standard deviant" instead of seeking to define a "standard normal." While there is only one way to qualify as a "standard deviant", a normal curve allows for richer diagnostic possibilities as there is any number of ways to deviate from the normal curve. It is our hope that the current research will allow norm-referenced decisions to be made that will expand the diagnostic possibilities available to clinicians. The Affinity 2.0 may have promising prognostic possibilities although further research is needed to determine the prognostic utility of this instrument. For instance, it seems possible that a young offender with a non-pedophilic presentation on the Affinity 2.0 may have less chance of re-offense than an individual with a profile that suggest the presence of pedophilic sexual interest.

Of final clinical significance is the possibility of creating local norms using the Affinity 2.0. The Affinity 2.0 is an inexpensive, non-invasive, easily administered instrument that requires relatively little operator instruction. Given this ease of administration, it seems plausible that local norms could be attained using this instrument, that would be more representative of a given subset of people than would a universal norm. Whereas universal norms may say relatively little about any specific population, local norms might create a normative curve far more representative of a specific sample of individuals. This might allow clinicians to make prognostic, diagnostic, and screening decisions using a normative curve that more closely represents the clients they work with in regards to important diversity factors such as ethnicity, age,

and educational attainment. David Glasgow and Justine Croxen (2003) administered the Affinity 2.0 to 27 adult, non-offending males in the United Kingdom. When compared with this sample, there is remarkable similarity between the two curves. Thus, in addition to the possibility of creating highly representative local norm samples using the Affinity 2.0, it also seems plausible that there may be a great deal of similarity between groups, even groups separated by a great distance. Further research needs to be conducted to examine the applicability of the local norms hypothesis to the Affinity 2.0, as well as testing the similarity or dissimilarity of normal curves obtained from areas that differ in culturally important ways.

Strengths

No study is without its strengths and weaknesses, and this study is no exception to that rule. One strength of this study is that it represents one of the few times that temporal stability has been examined in the context of viewing time measures of sexual interest. Most previous studies have examined only time-one administration of the VT instrument, thereby disallowing any inferences to be made concerning temporal stability. One strength of this study is that it takes into account that sexual interest is hypothesized to be a temporally stable construct, and should be measured thusly on a reliable instrument. One these grounds, we suggest that future research regarding viewing time measures of sexual interest should also includes some gauge of temporal stability.

Existing studies of viewing time measures of sexual interest have not produced norm-referenced data, thereby limiting their statistical strength and generalizability. Past research on VT measures of sexual interest has only reported ipsatized scores which do not allow for statements of “normality” and “abnormality” to be made. One strength of

this study is that it produces norm-referenced data, which allows for statements of normality to be made. Further, it provides a characteristic curve, from which an infinite number of deviations can be derived and examined. It is our hope that future examinations of viewing time measures of sexual interest will also include measures of temporal stability, an important consideration historically overlooked in similar research.

A third strength of this study is the large *n* of 120 test and 88 retest observations obtained. Previous studies have had limited generalizability and questionable robustness as a result of the small n's obtained by the researchers. The large number of participants obtained in our study makes this study stable and provides a large sample from which to make broader inferences. It is our hope that future research in this area will consider statistical issues of sample size when undertaking similar research.

A final strength of this study is that the data in their entirety are available to the researchers for whatever statistical manipulations are deemed necessary. This freedom allows new avenues to be pursued and permits researchers to critically examine the data in an open and uninhibited manner. As evidenced by a small number of our statistical procedures, analyses initially hypothesized to fruitful may prove useless under closer scrutiny. Conversely, full access to data also allows researchers to make improvements on existing statistical procedures and means of examining the viewing time data. For instance, our analysis of the temporal stability of the responses in this sample would not have been possible had we not been given full access to all of the raw data of our sample. Some extant measures of VT sexual interest do not allow access to participants' raw data, which we feel to be a detriment to both the researchers and the consumers of research on which the instrument was used. It is our hope that in the future, researchers and

consumers of research will be privy to all of the raw data as well as all of the statistical procedures undertaken as part of a given study.

Limitations

Research regarding something as private as sexual interest is likely to have some limitations. Strassberg and Lowe (1995) showed that volunteers for research involving sexual topics are generally more sexually experienced and hold more positive attitudes toward sex than the general population. While this bias is possible in our sample, it is unclear whether or not sex positive attitudes would have affected performance on the Affinity 2.0, as the true task was disguised to begin with. Additionally, the study was conducted at Brigham Young University, a school funded and closely aligned with The Church of Jesus Christ of Latter-day Saints (Mormons or LDS Church). The LDS Church requires BYU students to sign and adhere to a strict code of moral conduct (Honor Code) which includes rules regarding sexual behavior. Students may be dismissed from the school if they are found to be engaging in sexual (or other) behavior not in accordance with the Honor Code. Although efforts were made to assuage the fears of our participants regarding being reported to school authorities, it seems plausible that a minority of students allowed fear of school authorities to influence their test results. For example, three students marked that they had some homosexual experience (2 or 3 on the Kinsey Scale in this case), before ultimately declaring themselves entirely heterosexual (1 on the Kinsey Scale). While this could have been mere accident or coincidence, it is possible that BYU's strict code regarding sexual behavior influenced both participant selection and responses.

A second limitation is that our sample was largely young, well educated, and Caucasian. Only two of the participants in our study fell into the category of "non-traditional student," both of whom were only 45 years old. It is as yet unclear how an older sample would respond to similar administrations of the Affinity 2.0. All of the participants in the study were students at Brigham Young University (BYU) in Provo, Utah. Brigham Young University is a competitive school with above-average entrance requirements. Additionally, extra credit was offered to many participants involved in research classes, which may have furthered biased the participant selection toward students concerned with high achievement. Therefore, although no measure of intelligence was administered, it seems safe to assume that the majority of our sample was of above average intelligence.

Finally, a majority of our participants were Caucasian, thereby limiting the generalizability of the results in regards to some cultural variables. Currently, there is no data on multicultural responses to the Affinity 2.0. It seems probable that cultural conceptions of beauty play into both self-report and viewing time responses to the Affinity 2.0, thereby affecting outcomes. Furthermore, racist attitudes may affect participant ratings of slides of people of a given race (Glasgow, personal communication October, 2004). The slides included in the Affinity 2.0 depict a number of different racial groups in an effort to minimize cultural bias and measure the desired construct. However, until a thorough study of multicultural responses to the Affinity 2.0 is undertaken, we are uncertain what effect the relative cultural homogeneity of our participants has on our final results.

The number of participants originally proposed as part of the prospectus defense was 100 at both test and retest. While well over 100 men participated in the initial administration of the Affinity 2.0, the number that returned for a second administration was lower (88). This trend towards low participation surprised us, and ran contrary to our original hypothesis that men would be more willing than women to participate in research involving sexual interest. Harmon seemed to have little trouble reaching and exceeding her 100 participant threshold, while our study only reached 88 after two years of data collection. While it is still unclear what the reasons for this diminished return rate may be, a number of hypotheses exist. First, many of the participants felt cheated at having been promised "two movie tickets" upon realizing that the movie tickets in question were to the dollar movie theater (total cash value $2). Some members of the research team advertised the incentive as "two movie tickets," while others mentioned "two dollars off of a movie ticket" or "two tickets to the dollar movies." A number of participants felt deceived at the small monetary value of the incentive and may not have returned to show their displeasure. A second factor was that women in Kara Harmon's study seemed anecdotally to place more value on the extra credit offered in conjunction with completing both test and retest portions of the study. As women often outperform men in a collegiate setting, it may be safe to infer that grade incentives were more powerful for women than men in this case. Finally, a number of the participants reported feeling "weird" or "grossed out" at having to rate children in the context of a study on sexual interest. Administrators of the Affinity 2.0 received frequent complaints regarding participant discomfort with the task of rating children in the given context. It is possible that some participants did not return, as they were uncomfortable participating in a study

that involved sexuality and children. Although the study did not reach the proposed goal of 100 participants at both test and retest, the sample obtained is very robust and represents a vastly improved sample size over extant research in this area.

One final area of improvement for the Affinity 2.0 would be the addition of slides to the assessment. The Affinity 2.0 as currently constituted has seven slides for each of the eight stimulus categories. An increase in the number of slides per stimulus category would likely positively affect the reliability of the measure. The Affinity 2.5, currently in production, will include ten slides in each of the eight stimulus categories as a way of addressing this concern and increasing the overall reliability of the instrument, as noted by L. Fischer (personal communication, January 18, 2007).

Implications for Future Research

Our study sought to set strict parameters for inclusion, and then to demonstrate the characteristic sexual interest curve for that sample, as measured by the Affinity 2.0. We feel as though our study accurately portrays the temporal stability and the characteristic curve for normal heterosexual males as measured by the Affinity 2.0. However, as many individuals do not fall within the parameters of our study, there is a great deal of future research that may be undertaken, using our research as a referential base.

One such study would be to examine the normal (non-pedophilic) homosexual male response to the Affinity 2.0. Our findings suggest that the Affinity 2.0 can provide a reliable measurement of the sexual preferences of heterosexual males, as measured by viewing time. Preliminary results such as this study suggest that the Affinity 2.0 may reliably outline a normal homosexual response curve just as it was successful in outlining

the normal heterosexual response curve. Further research needs to be undertaken to examine the reliability and temporal stability of ascertaining such a characteristic curve.

Research is currently underway that examines the normal adolescent response to the Affinity 2.0. One advantage of an adolescent response curve is that it would allow comparisons to be made to the adult male and female curves currently in existence. Given that the onset of sexual deviance is often in adolescence, research in this area may also become instrumental to providing early detection and treatment options for adolescents and their parents. It remains to be seen, what similarities may exist between an adolescent sample of normal heterosexual individuals and our own sample.

A third area of potential research involves identified pedophiles with clear focal preferences (e.g. – consistent preference for small child female victims). Our study systematically excluded those with a history of pedophilia or pedophilic interests. Just as our study demonstrated that "normal" participants did not exhibit pedophilic interest, a study of pedophiles should hypothetically be able to pinpoint their specific predatory sexual interests. If the Affinity 2.0 is able to successfully demonstrate the specific pedophilic interest of known predators, it would become a valuable tool in monitoring client progress and treatment planning.

Finally, our study consisted largely of Caucasian participants from a largely homogenous religious and cultural background. It is as yet unknown how cultural variables such as perception of beauty, relation of skin tone to perceived beauty, and body size and perceived beauty would affect administrations of the Affinity 2.0. in a more ethnically diverse sample. Further, it is largely unknown what effect the religious

affiliation of the participants had on our results. Our research might serve as a valuable comparative base for other research focusing on samples with different sexual mores.

Conclusion

In summary, the goal of this study was to examine whether or not a characteristic viewing-time curve exists for normal, heterosexual males, and to determine whether or not that curve is temporally stable. Our findings suggest that a temporally stable viewing-time curve does exist; a finding that has considerable theoretical and practical utility. The establishment of such a curve allows norm-referenced judgments to be made and outlines a normal curve from which an infinite number of possible deviations can be examined. This study was unique among VT studies in that the researchers were given total access to the raw participant data. This uninhibited access allowed us to run chi-square goodness of fit tests that show potential in detecting dissimulation and for completing case-by-case comparisons of data. Although our research suggests that the Affinity 2.0 may be a promising instrument, further research is needed to determine the usefulness of the instrument with populations that differ in regards to sexual orientation, age, ethnicity and other diversity variables. While it is certainly has potential as a non-invasive, ethical assessment of sexual interest, further studies are necessary to determine whether the Affinity 2.0 truly provides an accurate measure and classification of sexual arousal and preference.

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. *Sex Abuse: Journal of Research and Treatment, 10*(2), 81-95.

Adams, G. (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. *Human Development, 20*(4), 217-239.

Barrett, P. (2002). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *The Journal of Managerial Psychology.* Retrieved January 30, 2004 from http://www.pbarrett.net/ jomp2003web.pdf

Berry, D., & McArthur, L. (1985, February). Some components and consequences of a babyface. *Journal of Personality and Social Psychology, 48*(2), 312-323.

Brigham Young University (n.d.). Church educational system honor code. Retrieved January 30, 2004 from http://campuslife.byu.edu/HONORCODE/ chaste.htm

Brown, M. (1979). Viewing time of pornography. *The Journal of Psychology, 102,* 83-95.

Brown, M., Amoroso, D.M., Ware, E.E., Pruesse, M., & Pilkey, D.W. (1973). Factors affecting viewing time of pornography. *The Journal of Social Psychology, 90,* 125-135.

Bruning, J.L., & Kintz, B.L. (1997). *Computational Handbook of Statistics* (4th ed.). New York: Addison-Wesley Education Publishers Inc.

Cattell, R. (1944, September). Psychological measurement: normative, ipsative, interactive. *Psychological Review, 51*(5), 292-303.

Cavior, N., & Lombardi, D. (1973, January). Developmental aspects of judgment of

physical attractiveness in children. *Developmental Psychology, 8*(1), 67-71.

Clemans, W.V. (1956). An analytical and empirical examination of some properties of ipsative measures. *Psychometric Monographs, 14.*

Cross, J., Cross, J., & Daly, J. (1971, December). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10*(6), 393-396.

Cross, J., & Cross, J. (1971, November). Age, sex, race, and the perception of facial beauty. *Developmental Psychology, 5*(3), 433-439.

Cunningham, M. (1986, May). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology, 50*(5), 925-935.

Dion, K. (1973, September). Young children's stereotyping of facial attractiveness. *Developmental Psychology, 9*(2), 183-188.

Fischer, L. (2000). The Abel screen: A non-intrusive alternative? In D.R. Laws, S.M. Hudson, & T. Wards (Eds.), *Remaking relapse prevention with sex offenders: A sourcebook.* Thousands Oaks, CA: Sage Publications.

Fischer, L., & Smith, G. (1999). Statistical adequate of the Abel assessment for interest in paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11*(3), 195-205.

Freund, K., Langevin, R., Cibiri, S., & Zajac, Y. (1973, February). Heterosexual aversion in homosexual males. *British Journal of Psychiatry, 122*(567), 163-169.

Glasgow, D. (2003). Affinity 2.0 [Software and Instruction], David Glasgow. Database Right.

Glasgow, D.V., Croxen, J., & Osborne, A. (2003). An assessment tool for investigating pedophile sexual interest using viewing time: An application of single case

methodology. *British Journal of Learning Disability, 31*, 96-102.

Gray, S. (1999, September). Outcomes of the Abel assessment and the penile plethysmography in a sample of sex offenders in outpatient treatment. In S. Johnson (Chair), *Current research comparing plethysmography with the Abel Assessment for Sexual Interest.* Proceedings from symposium at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, FL.

Harmon, K. (2006). *Non-pedophilic heterosexual female response to the Affinity 2.0.* Unpublished doctoral dissertation, Brigham Young University, Provo, Utah

Harris, G.T., Rice, M.E., Quinsey, V.L., & Chaplin, T.C. (1995). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34*(4), 389-394.

Horvath, T. (1981, February). Physical attractiveness: The influence of selected torso parameters. *Archives of Sexual Behavior*, *10*(1), 21-24.

In the interest of CDK, JLK, and BJK, S.W. 3rd (Amarillo, Texas, Appellate Court, 2002)

Kaufman, K.L., Rogers, D., & Daleiden, E. (1998, October). *Sexual interest in incarcerated adolescent offenders: A test-re-test analysis.* Poster session presented at the annual meeting of the Association for the Treatment of Sexual Abusers, Vancouver, BC.

Kinsey, A.C., Pomeroy, W.B., & Martin, C.E. (1998). *Sexual behavior in the human male.* Bloomington, IN: Indiana University Press.

Korthase, K., & Trenholme, I. (1982, June). Perceived age and perceived physical attractiveness. *Perceptual and Motor Skills*, *54*(3), 1251-1258.

Krueger, R., Bradford, J., & Glancy, G. (1998). Report from the Committee on Sex

Offenders: The Abel Assessment for Sexual Interest: A brief description. *Journal of the American Academy of Psychiatry and the Law*, *26*(2), 277-280.

Langlois, J., & Roggman, L. (1990, March). Attractive faces are only average. *Psychological Science*, *1*(2), 115-121.

Love, R.E., Sloan, L.R., & Schmidt, M.J. (1976). Viewing pornography and sex guilt: The priggish, the prudent, and the profligate. *Journal of Consulting and Clinical Psychology*, *44*, 624-629.

Mandell, R. (n.d.). Evaluating shorter versions of the Marlowe-Crown Social Desireability Scales. National Institutes of Health. Retrieved January 30, 2004, from http://www1.od.nih.gov/behaviorchange/measures/PDF/MC10_scoring.pdf

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. & Fernandez, Y. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review*, 20, 807-822.

Martin, B. (1964). Expression and inhibition of sex motive arousal in college males. *Journal of Abnormal and Social Psychology*, *68*, 307-312.

McClave, J.T. & Sincich, T. (2006). *Statistics* (10th ed.). New York: Prentice Hall.

Mueser, K., Grau, B., Sussman, S., & Rosen, A. (1984, February). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology*, *46*(2), 469-478.

Quinsey, V., & Chaplin, T. (1988, August). Preventing faking in phallometric

assessments of sexual preference. *Annals of the New York Academy of Sciences, 528*, 49-58.

Quinsey, V., & Chaplin, T. (1988, September). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence, 3*(3), 259-274.

Quinsey, V.L., Rice, M.E., Grant, T.H., & Reid, K.S. (1993). The phylogenetic and ontogenetic development of sexual age preferences in males: Conceptual and measurement issues. In H.E. Barbaree, W.L. Marshall, & S.M. Hudson (Eds.), *The Juvenile Sex Offender*. New York: Guilford.

Quinsey, V.L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology, 17*(5), 341-354.

Rosenweig, S. (1942). The photoscope as an object device for evaluating sexual interest. *Psychosomatic Medicine, 4*, 150-158.

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, 1*(3), 207-216.

Sneed, B.J. (2006). *The use of viewing time of non-erotic stimuli to differentiate between frequent pornography users and infrequent pornography users.* Unpublished doctoral dissertation, Brigham Young University, Provo, Utah.

Strahan, R., & Gerbasi, K.C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191-193.

Strassberg, D., & Lowe, K. (1995, August). Volunteer bias in sexuality research. *Archives of Sexual Behavior, 24*(4), 369-382.

Wright, L.W., & Adams, H.E. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment, 16* (3), 221-231.

Zamansky, H.S. (1956). A technique for assessing homosexual tendencies. *Journal of Personality, 24*, 436-448.

APPENDIX A

Demographics, Attitudes, and Sexual Interest Questionnaire

Demographics

1. Age: ____

2. Ethnicity: _____

3. Year in School (mark the one that applies)

___Freshman ___Sophomore
___Junior ___Senior
___Graduate Student

4. Marital Status

___Single ___Married
___Divorced ___Widowed

5. Did you hear about this research study through a psychology class?

___Yes ___No

Personal Attitudes

6. Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to your personality.

____ I never hesitate to go out of my way to help someone in trouble.
___ I have never intensely disliked someone.
___ There have been times when I was quite jealous of the good fortune of others.
___ I would never think of letting someone else be punished for my wrongdoings.
___ I sometimes feel resentful when I don't get my way.
___ There have been times when I felt like rebelling against people in authority even though I knew they were right.
___ I am always courteous, even to people who are disagreeable.
___ When I don't know something, I don't at all mind admitting it.
___ I can remember "playing sick" to get out of something.
___ I am sometimes irritated by people who ask favors of me.

Sexual Interest

7. I would describe my sexual preference as (please mark only one):

____ Exclusively heterosexual with no homosexual
___ Predominantly heterosexual with incidentally homosexual
____ Predominantly heterosexual with more than incidentally homosexual
___ Equally heterosexual and homosexual
___ Predominantly homosexual with more than incidentally heterosexual
___ Predominantly homosexual with only incidentally heterosexual
___ Exclusively homosexual with no heterosexual

APPENDIX B

Consent Form

Consent to be a Research Subject Participant # XXX

We would like to request your participation in a study designed to investigate a newly developed instrument that purports to measure sexual interest. You have been chosen for this study because you are 18 years old, and have no prior history of pedophilia. If you do not meet either of these requirements, you are ineligible and should withdraw at this point. As part of this research study, you will be asked to be asked to rate various line drawings and images of fully clothed people of both genders and of a variety of ages based upon their sexual attractiveness and unattractiveness to you. No pornographic images are used in this study.

The entire procedure should take no more than 30 minutes to complete. You will then be invited to repeat the rating process two to four weeks after the initial rating procedure. Afterwards you will be asked to fill out a questionnaire designed to gather simple demographic information, personal attitudes, and sexual interest. Since this is simply an assessment study rather than a treatment study, there are minimal risks to you. Upon full completion of this study you will be presented with two free movie passes as a token of appreciation for your participation. Although this study will yield no immediate personal benefits to you, it may yield long-term benefits to society in the future.

Your participation in this study is entirely voluntary and you are free to refuse to participate or stop at any time without penalty. Your grade or class standing will not be affected in any way of you decide to stop. All information will be number coded to insure your privacy. Only the researchers participating in this study will have access to your name, which will be kept strictly confidential. Your identity will not be revealed without your written consent and no identifying information will be made available to Brigham Young University's Honor Code Office.

If you have any questions, feel free to ask a participating researcher or contact us.

Daniel Crosby Lane Fischer, Ph.D.

(801) 224-7589 (801) 422-4200

cdc47@email.byu.edu

lane_fischer@byu.edu

If you have any questions or concerns that you do not feel comfortable asking the researcher, you may contact Dr. Shane Schulthies, IRB Chair, (801) 422-5490, 120 RB, shane_schulthies@byu.edu.

Please read the following paragraph, and, if you agree to participate, please sign below.

I agree to become a participant in the aforementioned study. I understand that any information about me obtained from this research study will be kept strictly confidential.

Signature_____ Date_____

Witness_____ Date_____

Please place your initials here to confirm that you have received a copy of this consent form. _____

APPENDIX C

Research Summary

Summary of Proposed Research

Specific Aims

The purpose of this research study is to assess the test-retest (temporal stability) of the responses of 100 normal heterosexual males to the Affinity 2.0, a viewing time measure of sexual interest.

Hypothesis

This study proposes to answer two major questions: 1.) Is there a characteristic pattern to normal heterosexual males viewing time responses to the Affinity 2.0? 2.) Is the Affinity 2.0 temporally stable?

Background and Significance

Viewing time has been utilized to distinguish between persons with high and low sexual interest (Rosenweig, 1942), varying levels of sex guilt (Love, Sloan & Schmidt, 1976), heterosexual and homosexual males and females (Zamansky, 1956; Brown, 1979; Quinsey et al., 1993, Quinsey et al., 1996, Wright & Adams, 1994), and male non-offenders and child molesters (Harris, 1996). Research has also shown that viewing time measures of sexual interest that incorporate a covert measure of viewing time are less transparent and thus less susceptible to dissimulation(Quinsey et al., 1993; Marshall, 1996). While male sexual interest can also be measured using phallometry (or penile plethysmography), this measure of sexual interest is ethically questionable, highly invasive, and lacks the potential for use with minors (Marshall, 1996). Viewing time measures of sexual interest such as the Affinity 2.0 avoid many of the problems associated with measuring male sexual interest by being non-invasive, using a covert

measure of viewing time, and non-pornographic stimuli (Glasgow, Osborne, & Croxen, 2003). These considerations make the Affinity 2.0 a good candidate for ethically measuring sexual interest in men of various ages.

Given that Affinity 2.0 is a newly standardized instrument we do not yet know what a normal heterosexual male response is to this instrument nor whether that response is temporally stable. If the normal heterosexual male response toAffinity 2.0 is a discernable and stable one this instrument could aid future research and clinical purposes.

Description of Subjects

Subjects will include 100 normal heterosexual males, all 18 years or older in age. Participants selected for this study will be sampled from the large population of students taking undergraduate psychology classes at Brigham Young University.

Confidentiality

The informed consent document that each potential participant must read and sign informs the participant as to the purpose of the study and expectations. Researchers will assure that all participants fully understand the informed consent documentation before they are asked to sign it. Also included is a section that discusses the confidentiality of the identity of the potential participants. It assures potential participants that all information pertaining to the participant will be kept confidential and that no names will be used in the study or reported to the Honor Code Office of Brigham Young University.

To protect the confidentiality of participants a master list of names will be associated with ID numbers. There will only be one master list that will kept in a locked file in a locked vault in the CPSE Lab. A research team of three members are the only persons who will have access to the master list. Performance results will be coded by

subject number and will be saved on the Affinity 2.0 hard drive and downloaded to research team's SPSS. Data will be deleted from Affinity 2.0 hard drive upon completion of analysis. Hard copies of results will also be stored in a locked file in a locked vault and will destroyed after five years.

Method or Procedures

Individuals attending psychology classes at Brigham Young University will be asked to participate via a short presentation by a researcher who will explain that the purpose of this study is to test a new device that claims to measure sexual interest. Potential participants will be informed that participation in the study would include looking at various line drawings and images of fully clothed people of both genders and of a variety of ages based on their sexual attractiveness and unattractiveness.

After signing the informed consent document, the Affinity 2.0 will be administered to participants on two separate occasions; at an initial assessment and then a second time two to four weeks later. Each visit will take no longer than 30 minutes. At the end of the second visit participants will complete the study by filling out a demographic questionnaire.

Data Analysis

Various independent and dependent statistical analyses will be utilized to describe the pattern of the viewing time data. Independent measures will include the calculating of sums, means and medians to compare inter-individual variation with regards to raw viewing time scores. Dependent, ipsative measures (category ranks, weighted ranks, standardized ipsative scores, and standardized raw scores) will be employed to compare intra-individual differences in viewing time. In addition, a "goodness of fit" χ^2 analysis

will be conducted to compare test and retest results to determine whether the normal heterosexual female response to the Affinity 2.0 is a temporally stable one.

Risks

Given that sexual interest is such a sensitive topic, the administration of the Affinity 2.0 and the questionnaire could cause some psychological discomfort. A potential breach of confidentiality regarding "sensitive information" is also possible.

Benefits

Participation in this study will yield no immediate personal benefits to the participant. However, the results of this study may yield long-term benefits to society in the future as it may allow for an enhanced psychometric assessment of normal and deviant sexual interest.

Compensation

Upon full completion of this study each participant will be presented with two free movie passes as a token of appreciation for participation.