2007-03-22

Stability of a Normal Heterosexual Female Response to Affinity 2.0

Kara L. Harmon
Brigham Young University - Provo

THE TEMPORAL STABILITY OF A NORMAL HETEROSEXUAL FEMALE RESPONSE TO AFFINITY 2.0

by

Kara Harmon

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

December 2006

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Kara Harmon

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	Ellie L. Young
Date	Steve A. Smith
Date	Melissa Allen Heath
Date	Aaron P. Jackson

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Kara Harmon 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

Barbara Culatta
Associate Dean
David O. McKay School of Education

ABSTRACT

THE TEMPORAL STABILITY OF A NORMAL HETEROSEXUAL FEMALE RESPONSE TO AFFINITY 2.0

Kara Harmon

Department of Counseling Psychology and Special Education

Doctor of Philosophy

The purpose of the current study was to evaluate the temporal stability of a normal heterosexual female response to the Affinity 2.0, a newly standardized viewing time (VT) instrument that purports to measure sexual interest. Participants were 120 female undergraduate and graduate students from a private university (mean age = 21.67 years) who met inclusion criteria of non-pedophilic interest/history, identified as “Exclusively Heterosexual” on the Kinsey Scale, and who took the Affinity 2.0 both at test and at retest (approximately two weeks later). Participants also filled out a questionnaire following the retest composed of demographic questions, a shortened version of the Marlowe-Crowne Desirability Scale-10 [M-C 2(10)], and declared their sexual interest on the Kinsey Scale.

Pearson correlation coefficients (PPMCC), Spearman’s Rho correlation coefficients, and a Chi-Square Goodness of Fit Test were all utilized to assess the temporal stability of the sample's response to Affinity 2.0. All PPMCC and Spearman's Rho correlations for VT were statistically significant at the $p < .01$ level; while practical significance for PPMCC could only be considered moderate as best (range of r =.28 - .63), Spearman's Rho correlations (range of r =.41 - .65) were stronger and approached practical significance. The existence of an overall distinct and discernable VT response pattern was apparent as the sample demonstrated identical categorical VT preferences both test and retest. These results were consistent with results from previous VT studies (Quinsey, Rice, Grant & Reid, 1993; Wright & Adams, 1994; Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996). In addition, this VT pattern was found to be stable from test to retest as assessed by a Chi-Square Goodness of Fit Test. An unexpected finding was the presence of elevated VT for adult and juvenile female images. Implications, limitations, and future studies are discussed.

ACKNOWLEDGMENTS

I want to thank my chair, Dr. Lane Fischer, for his continuous assistance and support throughout this project. I would also like to thank the members of my committee: Dr. Steve Smith, Dr. Ellie Young, Dr. Melissa Heath, and Dr. Aaron Jackson. Their interest and support in this project as well as the contributions each have made to my continuing professional development are greatly appreciated. I would also like to thank all of the members of the Affinity team who helped to gather data and contributed their time and effort to help make the completion of this project a reality.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

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 would be invaluable in that it would enable researchers and clinicians to track sexual interest throughout the lifespan, to aid these professionals as a screening, diagnostic and/or prognostic device for identifying deviancy, as well as serving as a tool for monitoring the success of ongoing therapy and assessing treatment outcomes.

Currently the four methods for measuring and classifying sexual interest include the clinical interview, self-report measures, measurements of genital response (i.e., penile or vaginal plethysmography), and viewing time (Quinsey, Rice, Grant & Reid, 1993). While all of these measures may reveal important information regarding sexual interest, not all are effective/accurate or appropriate for use in assessing the sexual interest of females.

Despite the fact that the clinical interview is frequently used as an effective means for assessing sexual interest, it is a much maligned process due to the fact that detection of dissimulation is a daunting, if not impossible, task (Marshall, 1996). The subjective nature of the clinical interview can be a downfall as well. An apprehensiveness or reluctance to talk about sexual matters, misinterpretation of the question asked, or a desire to sound "normal" can also compromise the validity of responses (Quinsey et al., 1993). In fact, after conducting a study with sexually dysfunctional women who were placed in a comprehensive sex therapy program in an attempt to increase sexual arousal, Wincze, Hoon and Hoon (1978) concluded that data obtained during interviews was highly misleading.

While self-report measures often provide more objective data, due to the limitations of transparency (Marshall, 1996) and potential for distortion (Wright & Adams, 1994), they are susceptible to dissimulation as well. Misreport due to shyness, confusion, or embarrassment is also possible (Quinsey et al., 1993). Singer (1984) and Heiman (1980) have also cited that, while men's subjective reports of their arousal have demonstrated consistency with physiological measures, this is not always true for women's self-reports—e.g., women may report arousal at low physiological levels and/or fail to report arousal even at maximum physiological response. Thus, self-report is not always a reliable indicator of sexual interest for females.

A seemingly more objective, yet intrusive, means for assessing sexual interest is that of measuring genital response through penile or vaginal plethysmography (wherein penile tumescence or vaginal pulse and blood flow are measured during sexual arousal). While these measures may be more resistant to distortion, they are invasive and typically require sexually explicit material in order to generate genital response. This requirement makes the use of such instruments with adolescents or children ethically questionable (Marshall, 1996). In addition, Marshall and Fernandez (in press) have also identified problems with the psychometric soundness of this method citing potential problems with standardization, temporal stability, criterion validity, data formats, and internal consistency. Hoon (1984) has declared that the physiologic measures of female sexual response, taken alone, are likely to be misleading as orgasmic capability and sexual arousal may be totally dissociated in women (Wince, Hoon & Hoon,1978). This type of assessment seems especially problematic for women considering that numerous other researchers have also noted the lack of correlation between subjective and physiological reports (Benson, 2003; Heiman, 1980; Hoon, 1984). Thus, for practical, ethical, and empirical reasons, plethysmography is not reliable for assessing the sexual interest of women.

The growing empirical evidence for sustained visual attention suggests it may be a promising avenue for detecting sexual interest in females. It has the advantage of being non-intrusive, of not necessarily requiring sexually explicit materials, and being covert (participants, unless informed, are not aware that their sexual interest is being assessed independently of their self-reports). Studies utilizing this method have been shown to successfully discriminate between groups of homosexual and heterosexual males and females, child molesters and non-offending males, high and low sex guilt groups, and sexually interested and uninterested patients (Harris, Rice, Quinsey & Chaplin, 1996; Quinsey, Ketsetzis, Earls & Karamanoukian, 1996; Wright & Adams, 1994; Quinsey et al., 1993; Love, Sloan & Schmidt, 1976; Rosenzwieg, 1942). All studies found significant correlations between sexual preference and viewing time.

While a majority of the studies were conducted solely with male participants, it is important to note that the studies that included female participants obtained comparable results. Wright and Adams (1994) discovered that females, like males, looked significantly longer at slides of the preferred sex stimuli. In fact, Quinsey et al. (1996) found that, while males did look longer at the slides of their preferred sex, the difference in time was not statistically significant from that of female viewers. Both studies found that women looked the longest at slides of the adult preferred sex stimuli with viewing time decreasing with the age of the model. So, while the patterns that characterize male participant viewing times are often more elevated than those of female participants when viewing slides of preferred adult sex stimuli, female participants also demonstrate predictable and discernable patterns of viewing time that indicate sexual interest. Thus, viewing time appears to be a respectable emerging indicator of sexual interest for women as well as for men.

The Abel Assessment for Sexual Interest (AASI) is an instrument that claims to measure sustained visual attention as an indicator of sexual interest (Abel, Huffman, Wargerg & Holland, 1998). However, reviews of the instrument reveal psychometric concerns regarding the instrument's data format, normative base, temporal stability, validity, internal consistency (Fischer & Smith, 1999), deficits in test-retest reliability (Smith & Fischer, 1999; Kaufman, Rogers & Daleiden, 1998), and the potential for dissimulation (Gray, 1999). Also, due to the fact that AASI z-scores are ipsative, it precludes the opportunity for researchers or clinicians to compare scores across individuals or groups. The usefulness of this instrument for clinical and research purposes is further compromised by the fact that users of the instrument are denied access to the raw data results (Fisher & Smith, 1999). Hence, due to psychometric concerns, the inability to compare groups or individual scores, and the inaccessibility of data to users of the instrument, the AASI is unable to effectively serve clinicians and researchers in their pursuit to answer normative questions regarding male and female sexual interest.

The Affinity 2.0 is a computer based assessment of sexual interest that plots covertly measured viewing time against self-report measures to create an individual profile of relative sexual interest by gender and age (Glasgow, Obborne & Croxen, 2003).

The Affinity 2.0 has circumvented many of the problems that obstruct the efficacy of the AASI in that the resulting test data are directly accessible to the user. This is advantageous in that the user can create norm-referenced scores from this data. The Affinity 2.0 also has the advantage of utilizing clothed models in non-suggestive settings. Thus, the Affinity 2.0, with its readily accessible data that can be norm-referenced and potential for use with a variety of populations, has emerged as a promising instrument for effectively assessing male and female sexual interest for both research and clinical purposes.

Statement of Problem

Given that Affinity 2.0 is a newly standardized instrument we do not yet know what a normal heterosexual female response is to this instrument nor whether that response is temporally stable.

Statement of Purpose

The purpose of this study is to determine a normal heterosexual female response to Affinity 2.0 as well as to determine whether the response is temporally stable.

CHAPTER II

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). Sexual arousal and preference/interest are important phenomenon to understand as the implications and effects are far reaching. Just as normal sexuality can enhance relationships and well being, the presence of deviancy (e.g. pedophilia, voyeurism, exhibitionism, etc.) can have severely detrimental consequences for individuals as well as for society as a whole. Hence, the discovery of a reliable instrument would be invaluable in that it would enable researchers and clinicians to track sexual interest throughout the lifespan, to aid these professionals as a screening, diagnostic and/or prognostic device for identifying deviancy, as well as serving as a tool for monitoring the success of ongoing therapy and assessing treatment outcomes.

Clarification of Terminology—Sexual Interest, Attraction, and Arousal

Given the prevalence of sexual phenomena, a myriad of terms have been used in its discussion. Due to this reality, a clarification of terminology is helpful. The terms most frequently utilized to discuss sexuality in the literature include *sexual interest*, *sexual attraction*, and *sexual arousal*. At times these terms have been used synonymously and interchangeably while at other times these terms are utilized in a more discrete manner. As there currently exist no clear and distinct demarcations to definitively differentiate between sexual interest, attraction and arousal, a model of sexuality can be helpful in conceptualizing the relationships and differences that exist. One such model is the social cognition model.

According to social cognition model, humans tend to think in a categorical manner about others (Macrae & Brodenhausen, 2000). Just as humans tend to develop categorical schemas for individuals based on race, occupation, disability, etc., it also stands to reason that these same categorical schemas/stereotypes likely exist for preferred sexual partners. Sexuality, then, could be conceived to exist on a continuum with interest being a precursor to attraction, which, in turn, is a prerequisite for arousal. Thus, within this framework, sexual interest would be categorized by a generalized stereotype of a sexually preferred individual (i.e., the individual being the preferred sex and age). Once the individual was established as an exemplar of the general preferred sexual stereotype, then one could begin to ascertain how sexually attractive the individual is. This evaluation of sexual attractiveness is likely to be based on a more specified prototype (e.g., preferred hair color, eye color, or skin color, a favored body build or height, etc.). If the individual being considered is taken to be attractive, then further cognitive processing (e.g., fantasy) is likely to lead to physiologic sexual arousal. Hence, in this model, sexual interest and attraction are considered to be discrete, yet complimentary, domains that serve as prerequisites for sexual arousal.

Four Extant Methods for Measuring and Classifying Sexual Interest

Currently the four methods for measuring and classifying sexual interest include the clinical interview, self-report measures, measurements of genital response (i.e., penile or vaginal plethysmography), and viewing time (Quinsey et al., 1993). This review of the literature will explore how appropriate and effective each of these types of assessment have proven to be with a female population.

Clinical interview. Marshall (1996) pointed out that the clinical interview is a much maligned process due to the fact that detection of dissimulation in interviews is a daunting, if not impossible, task. The subjective nature of the clinical interview can be a downfall as well. An apprehensiveness or reluctance to talk about sexual matters, misinterpretation of the question asked, or a desire to sound “normal” can also compromise the validity of responses (Quinsey et al., 1993). In fact, Wincze, Hoon and Hoon (1978) concluded that “outcome and follow-up data obtained during interviews may be highly misleading” (p. 48). Their study included women who reported low sexual arousal despite satisfactory marriages and the absence of anxiety. At the initial assessment of these women it became clear that the information regarding sexual arousal dysfunction obtained in the interview was accurate as all of the patient’s SAI (Sexual Arousal Inventory) scores ranged between the 5th and 25th percentile. Participants then took part in a comprehensive sex therapy program to see if they could increase their capacity for sexual arousal.

When the participants were again evaluated following therapy, researchers were unable to detect any objective evidence that the women’s capacity for sexual arousal had increased. Nonetheless, when the women were interviewed post-therapy and again at the follow-up two years later, all gave glowing reviews regarding their positive experience with their therapist, the increased understanding and knowledge they had gained regarding sex, the improved status of their sexual relationships, and their increased capacity for sexual arousal. In fact, one woman who had reported that she had been inorgasmic her entire life claimed at the follow-up to be orgasmic.

Researchers were doubtful of these claims, particularly the latter, as none of the women had manifested any clinically significant changes in their physiologic capacity for sexual arousal. Researchers attributed their results to social conditions rather than therapeutic efficacy, hypothesizing that the social expectations and pressures of being in therapy resulted in the participants reporting positive improvement and outcomes.

Thus, while some of the information gained in clinical interviews may be accurate and verifiable (like the women's initial claims of sexual arousal dysfunction), the accuracy of other information (like the claimed therapeutic benefits) proffered in clinical interviews may not be. It also becomes clear that information obtained from clinical interviews may not only potentially be distorted by dissimulation, embarrassment, or confusion, but may be subject to societal pressures as well.

Self-report measures. A major criticism of self-report measures is their vulnerability to distortions that can arise from an individual's response set, biases and/or motivation (Wright & Adams, 1994). Marshall (1996), while commending self-report measures for providing more objective data, also highlights the problematic transparency of such measures. Quinsey (1993) also points out that shyness with regards to talking about sexual matters, an inability to comprehend the questions, and/or the concern for one's responses to sound "normal" are also hindrances of the self-report method.

In addition, there are unique challenges that arise with regards to women's self report of sexual arousal. For instance, Singer (1984) and Heiman (1980) have cited that, while men's subjective reports of their arousal have demonstrated consistency with physiological measures, this is not always true for women's self-reports—e.g., women may report arousal at low physiological levels and/or fail to report arousal even at maximum physiological response. Self-report measures in the Wincze, Hoon and Hoon (1978) study were unable to be verified by objective physiological measurement. Heiman (1980) found that when participants subjectively reported negative affect, i.e. “disgust,” in response to an erotic tape, this was, nonetheless, correlated positively with physiological sexual arousal. Thus, there is much evidence that self-report is not always a reliable indicator of sexual interest for females.

Singer (1984) offered three hypotheses that might help explain the discrepancy between subjective and objective reports of sexual arousal in women. First, he suggested that women’s genital response may be just as distinguishable and strong as men’s but that subjective biases (e.g., the appropriateness of or confidence in revealing such disclosures, mislabeling, etc.) hinder/impede accurate report. Second, he proposed that women’s sensation of arousal may not be as strong as men’s and, thus, is more difficult to detect and report. Last, he proferred that women’s sexual arousal may be altogether qualitatively different from men’s and, thus, is more difficult to report accurately.

Genital response. Meston is the most recent researcher to confirm that “low, if any correlations” exist between the subjective reports and physiological arousal in females (2003; personal communication, January 30, 2004). In fact, vaginal plethysmography has proven to be a particularly erroneous assessment of sexual interest or arousal in women who may fail to report arousal even at maximum physiological response or report arousal at low physiological levels (Singer, 1984; Hoon, 1984; Wincze, Hoon & Hoon, 1978; Heiman, 1980). Results from the aforementioned study by Wincze, Hoon and Hoon (1978) led the researchers to affirm that orgasmic capability and sexual arousal may be totally dissociated in women. Thus, Hoon (1984) warned that the “physiologic measures of female sexual response, taken alone, may… be misleading” due to the fact that they “may not reflect the subjective experience of women” (p. 778).

This admonition was supported by the results of a study conducted by Heiman (1980). Fifty-five women participated in the study; 27 were married, 28 were single, and all were heterosexual. All the women participated in three different activities: viewing an erotic film, listening to an erotic audiotape, and participating in a self-generated erotic fantasy. A vaginal photoplethysmograph was used to monitor and measure the physiological sexual arousal (vaginal pulse amplitude—VPA) of each woman during each of the conditions. In addition, the participants rated their subjective arousal to each sexual stimulus on a five-point scale, as well the degree to which they experienced certain positive and negative affective states (enjoyment, guilt, disgust, interest, embarrassment).

Heiman (1980) was surprised to note the lack of correlation that existed between subjective and physiological measures of arousal. While the lower third of the sample showed negative correlations between VPA and reports of sexual arousal, the upper third manifested significant positive correlations. Further, half of the sample (the married women) did not yield any significant correlation between sexual arousal ratings and VPA at all. In addition, it was also discovered that VPA was also unrelated to affective responses. Even the results obtained from each woman during the two different sessions were not correlated.

Thus, while Singer (1984) cited that (with the exception at low levels of tumescence) the subjective reports of arousal for men correspond well with physiological measures (i.e. penile plethysmography), research has verified that this is not the case for women (Benson, 2003; Heiman, 1980; Wincze, Hoon & Hoon, 1978). Hence, it would be advantageous to find a more reliable objective method to assess female sexual interest.

V*iewing time as a discriminate of sexual interest—Underlying theory*. The last method for assessing sexual interest is that of viewing time (VT), or sustained visual attention. There are two different theories in particular that seem to provide support for the utilization of VT as a discriminate of sexual interest.

The social cognition model, discussed previously, seems to fit well with the VT phenomenon. In fact, sustained VT may be viewed as a special case of social cognition. For instance, when the subject of interest does not fit the generalized stereotype (i.e., the individual is too old/young, a member of the non-preferred sex, etc.) then the visual attention given to the individual is likely to be minimal. However, when the subject of interest is an exemplar of this initial generalized stereotype that subject is, in turn, further considered to determine how well he/she satisfies the more specified prototype. If the subject of interest successfully fits the prototype preferences, further cognitive activity (i.e., fantasy) is likely to take place. This extended cognitive activity and processing of resulting related associations (e.g., memories, fantasy, etc.) requires more time and, thus, is a plausible explanation for prolonged gazing at sexually preferred stimuli.

Another theory that supports the VT phenomenon is that of Singer (1984) who suggested that one of the components in the trichotomy of sexual arousal is the aesthetic response. The other two proposed components are approach and genital response. Singer describes the aesthetic response as an hedonic feeling in response to a sexual stimulus, i.e., the sight of an attractive face. He goes on to explain that a person displaying such a response would "make efforts to keep the object in view by means of eye movements or head turning” (p. 233). In other words, an individual would spend longer looking at stimuli he/she considers sexually attractive to him/her than to stimuli that are not. This is the basis for instruments that measure viewing time as an indicator of sexual interest. So, what has been discovered with regards to viewing time as an indicator of sexual interest in general?

Early studies of viewing time

Early studies of viewing time have revealed that it is a successful method for differentiating between various groups of people. Rosenzweig conducted one of the pioneer studies of viewing time in 1942. In a study of hospitalized male psychiatric patients he found that those patients who were highly interested in sexual topics spent longer looking at photographs that depicted sexual stimuli than did patients in the low interest group.

In 1956, Zamansky utilized non-erotic stimuli (pictures of clothed male and female models and landscape scenes) to successfully distinguish between homosexual and heterosexual males. Participants were presented with a pair of pictures (male/female, female/neutral, male/neutral, or neutral/neutral stimuli) and judged which of the pictures covered more surface area. Unbeknownst to the participants, all but the neutral landscape pictures were identical in area. During this task Zamansky used an instrument to track the eye movements of the participants and discovered that the heterosexual males spent more time looking at pictures of females than males whereas the homosexual participants spent longer looking at males than they did pictures of females.

Twenty years later, Love, Sloan and Schmidt (1976) found that viewing time was also able to differentiate between college males who were low, moderate, or high in sex-guilt. Through gauging the amount of time spent viewing increasingly explicit sexual stimuli slides it was discovered that researchers could readily distinguish between the three groups. Low sex-guilt group members manifested an increasing linear pattern of viewing time, moderate sex-guilt group participants demonstrated a curvilinear pattern of viewing time, and high-sex guilt group members did not show any significant increase in viewing time. In addition, the average time members of the high-sex guilt group spent looking at any explicit photo was significantly less than the average time for the other two groups.

Hence, according to early research findings, viewing time, measured surreptitiously as participants are performing a simultaneous task involving the preferred stimuli of interest, has been successfully utilized to distinguish between high and low sexual interest groups, heterosexual and homosexual men, and diverse sex guilt groups (Love, Sloan & Schimdt, 1976; Rosenzweig, 1942; Zamansky, 1956).While these findings are interesting, one might question what relevance, if any, the outcomes of these studies have for women given that all of these early studies were conducted utilizing only male participants.

Sustained Visual Attention to Visual Stimuli—Stereotypes vs. Research Findings

While it is not a new notion that “visual stimuli are a more potent source of sexual arousal for men than they are for women” (Symons, 1979, p. 146), it should not be assumed that sustained attention to visual stimuli is not a valid measure for assessing female sexual interest. While one may expect women to “report more arousal to romantic fantasy than to visual erotica” (Singer, 1984, p. 235), it should not be assumed that this is the case.

In 1977, Heiman conducted a study with 59 female and 39 male undergraduate students. Participants were exposed to a series of four different audio tapes that varied in their erotic and romantic content. It was discovered that both women and men were sexually aroused by explicitly erotic content and that women's responses were not enhanced by the presence of romantic themes. She also found that women did not demonstrate greater sexual arousal to a romantic tape than they did to a control tape.

In 1979, Brown made another unexpected discovery with regards to female viewing time and sexually explicit visual stimuli. Thirty male and 30 female undergraduate students were recruited for a study to test whether exposure to increasingly hard-core pornographic images would evoke avoidance reactions in participants. Each participant was exposed to 21 slides with varying sexual themes (male/female nudes, male/female masturbation, heterosexual intercourse and oral-genital activity, heterosexual group sex, lesbianism, homosexuality, and male transvetism). Participants were informed that they could view the slides for as long as they wished as they would be evaluating the photographs immediately after viewing them. Slides were rated on seven-point semantic differential scales according to their evaluative and activity factors, as well as how arousing the image was. The viewing times for the slides were recorded as participants completed their task.

Though the viewing times of males were significantly longer for all but two of the slides, Brown (1979) was surprised to find that, while males demonstrated a curvilinear pattern of viewing, the females did not. In fact, viewing times for females gradually increased throughout the exercise, even when looking at the most blatantly pornographic slides.

Clearly, the assumption that romantic or relational themes are necessary for women to be attentive to visual stimuli is erroneous (Hoon, 1978; Heiman, 1977; Brown, 1979). On the contrary, the research supports Hoon's (1984) assertion that "women respond to direct representations of erotic activity much the same way that men do" (p.778). So, though males may be expected to be more responsive to visual sexual stimuli (Symons, 1979; Bailey, Gaulin, Agyei & Gladue, 1994), research has shown the response of men and women to be more similar than previously believed.

Viewing Time Studies Conducted with Females

In 1993, Quinsey, Rice, Harris and Reid attempted to assess the sexual interest of 15 male and 15 female heterosexual participants by exposing each participant to 31 slides that exhibited nude males and females in four different age categories: infant, children, pubescents, and adults. Participants were instructed to obtain a clear view of the model in each slide before advancing to the next because afterwards he/she would be asked to make some ratings about each. After viewing all of the slides for the first time, participants were then directed to rate on a scale of 0 to 100 how physically attractive they felt the model was to people in general as well as how sexually attractive the model was to him/her personally.

While viewing time was not found to be a strong measure of sexual preference in this study, males and females did exhibit different viewing time patterns that allowed researchers to distinguish between the two groups. Males viewed slides of both adult and pubescent females slides for longer periods of time than all categories of non-preferred slides. Females tended to look at both preferred and non-preferred slides for relatively equal amounts of time, although there was little difference in viewing time across age and gender conditions, Quinsey et al. (1993) found considerable correlations between viewing time and sexual attractiveness ratings for both males and females.

Three years later, Quinsey, Ketsetzis, Earls and Karamanoukian (1996) conducted a similar study. They hoped to replicate their previous finding of the substantial correlations that existed for both males and females between viewing time and sexual attractiveness ratings (Quinsey et al., 1993). They also sought to confirm four new hypotheses with regards to VT: (1) that both male and female participants would view slides of young adults of the opposite sex the longest and adults of the same sex and prepubescent children for the shorter amounts of time; (2) that males would have a higher correlation between sexual attractiveness and viewing time than would females; (3) that males would look at slides of pubescent females longer than women would look at slides or pubescent males; (4) that males would look longer at slides of adult females than females would look at slides of adult males.

This study differed slightly from the Quinsey et al. study conducted in 1993 in that the number of heterosexual participants was increased to 24 male and 24 females, and the 36 slides of male and female models included only three age categories: adult, pubescent and child. Again, all models were nude and in full frontal view, and none appeared flirtatious. Just as before, participants viewed the slides through the use of a projector with two buttons—one button advanced the slide while the second button illuminated it. Participants were instructed to carefully view each slide as some general questions regarding the slides would be asked after viewing them. After participants had viewed the slides for the first time they repeated the procedure, but this time participants were also asked to rate each slide on a seven-point Likert scale with regards to how sexually attractive the model appeared. The amount of time that each slide was illuminated by the participant was recorded in milliseconds.

After computing an average for every participant for the six slides within each age and sex category, researchers found that both males and females looked at adults of the preferred sex longer with viewing times decreasing with the age of the models. Both rated the slides of the adult member of the opposite sex the highest. Researchers were able to confirm all of their hypotheses but one; they were surprised to find that, while male participants did illuminate the slides of their preferred sex longer, the difference in time was not statistically significant from that of females.

Wright and Adams (1994) used sexually explicit material to differentiate between the sexual preferences of four groups: heterosexual females, homosexual females, heterosexual males, and homosexual males. They recruited 20 participants from each of the four groups of interest. Participants were then exposed to 60 different slides, which contained pictures of *Playgirl* models, *Playboy* models, or landscape scenes. A white dot was superimposed in one of five locations on the slide (upper-right corner, upper-left corner, lower-right corner, lower-left corner, or middle). While viewing each slide each participant was given the choice reaction time task to locate the white dot as quickly as possible and then to press the corresponding button on the hand-held box given to him or her. The time spent and the accuracy of the report given by the participant was recorded in the computer after each slide was viewed.

Wright and Adams (1994) found that all participants from the four groups demonstrated significantly longer latencies in choice reaction time on the slides that depicted a preferred sexual stimulus than on the slides that contained a non-preferred or neutral stimuli. Though the patterns for women were less elevated than those for the men, they were distinct and discernable. They also discovered that they could readily distinguish not only both female groups from both male groups, but heterosexual females from the homosexual females as well.

Thus, all three of these studies (Quinsey et al., 1993; Quinsey et al., 1996; Wright & Adams, 1994) found that substantial correlations existed for both males and females between viewing time and preferred sexual stimuli. The Quinsey et al. (1996) and Wright and Adams (1994) studies found women to look the longest at slides of the adult preferred sex stimuli with viewing time decreasing with the age of the model. In these studies the VT patterns of women, though less elevated than those of men, were unique and distinguishable enough for researchers to easily differentiate between heterosexual females and males as well as heterosexual and homosexual females.

Conclusions/Implications with Regards to VT Studies

In reviewing the research it has been found that VT has proven to be successful in distinguishing between persons with high and low sexual interest (Rosenzweig, 1942), homosexual and heterosexual males (Zamansky, 1956; Wright & Adams, 1994), varying levels of sex guilt (Love, Sloan & Schmidt, 1976), heterosexual males and females (Brown, 1979; Quinsey et al., 1993; Quinsey et al., 1996; Wright & Adams, 1994), homosexual females and homosexual males (Wright & Adams, 1994), heterosexual females and homosexual males (Wright & Adams, 1994), and homosexual females and heterosexual males (Wright & Adams, 1994).

Research has also revealed that males and females are not so disparate in their response to visual stimuli. Though VT are generally greater for men, the Quinsey et al.

(1996) study did not show this difference to be significantly greater than that of women's VT. While VT patterns demonstrated by women are not as elevated, they have proven to be just as distinguishable and discernable as VT patterns for men.

In 1995, Harris, Rice Quinsey, and Chaplin were able to utilize measures of VT to differentiate between non-offending males and offending child molesters. Though no like studies have been conducted with female participants, it seems logical that, if heterosexual and homosexual females manifest distinct and discernable VT patterns (Wright & Adams, 1994), then different viewing patterns would be apparent for different female sub-populations (deviant, dysfunctional, etc.), just as they have with males.

Two VT Instruments—The Abel Assessment for Sexual Interest and The Affinity 2.0

Given that VT is less susceptible to dissimulation than are self-report measures or clinical interviews (as data are measured covertly) and research has revealed VT to be more reliable indicator of female sexual interest than genital response instruments, it seems evident that surreptitiously measured viewing time is the best "candidate for…an ethically sound, unobtrusive measure of sexual preference" for women (Quinsey, 1993, p. 159). The next logical step then, is to find a reliable, temporally stable VT instrument that can aid both researchers and clinicians in their quest to accurately classify and measure sexual preference and arousal. There are two such instruments that can be evaluated on their potential to accomplish this task: The Abel Assessment for Sexual Interest (Abel, Huffman, Wargerg & Holland, 1998) and the Affinity 2.0 (Glasgow, Osborne & Croxen, 2003).

The Abel Assessment for Sexual Interest (AASI). The AASI is an instrument that claims to measure sustained visual attention as an indicator of sexual interest through both self-report and surreptitiously measured VT to photographs of various stimuli. This instrument has been marketed to serve screening, diagnostic, and prognostic functions (Fischer, 2000). However, upon review, psychometric concerns regarding the instrument's data format, normative base, temporal stability, validity, reliability (Fischer & Smith, 1999; Fischer, 2000), deficits in test-retest reliability (Smith & Fischer, 1999; Kaufman, Rogers & Daleiden, 1998) and the potential for dissimulation (Gray, 1999) have arisen. It is also important to note that, instead of relying on and measuring between-group variation (as all of the other research studies pertaining to VT have), the AASI focuses and reports only on intra-individual variance. Thus, in the resulting AASI ipsative scores there is no commensurability across participants or groups. As a result, the AASI precludes comparison between individuals or groups and, in turn, is unable to answer normative questions (Fischer & Morgan, 2006).

In addition to the questionable technical adequacy and subsequent scores that are yielded from the AASI, the usefulness of this instrument for clinical and research purposes is further compromised in that AASI researchers refuse to release the raw data to any user of the instrument (Fischer & Smith, 1999). Consequently, clinicians and researchers are wholly dependent upon AASI researchers for the interpretation of data as they are denied any direct access to resulting data scores. Considering the questionable psychometric foundation of the AASI, this is a particularly discomfiting predicament.

Hence, due to psychometric concerns, the inability to compare groups or individual scores, and the inaccessibility of data to users of the instrument, it can be safely concluded that the AASI is unable to effectively serve clinicians and researchers in their pursuit to answer normative questions regarding male and female sexual interest.

The Affinity 2.0. The Affinity 2.0 is a computer based assessment of sexual interest that plots covertly measured VT against self-report measures to create an individual profile of relative sexual interest by gender and age (Glasgow, Osborne & Croxen, 2003). Presently, this instrument is in its second revision. While Affinity was designed to assess the sexual interest of males with a mild learning disability, the current version is licensed for use as a clinical assessment tool for learning disabled adult male offenders and adult male offenders. Glasgow (2003) has also approved the use of the Affinity 2.0 to be used for research and evaluation purposes with adult male non-offenders, juvenile male offenders, and female sex offenders.

The Affinity 2.0 has circumvented many of the problems that obstruct the efficacy of the AASI. Firstly, the resulting test data from Affinity 2.0 are directly accessible to the user. This is advantageous in that it permits the user to know precisely where the scores came from and to discern what those scores mean. Hence, resulting raw scores are readily available for both clinical and research purposes. They can be utilized for normative study through the creation of norm-referenced scores. Commensurability allows comparison across individuals as well as groups.

It is important to note that a common limitation of VT instruments has been the use of sexually explicit visual stimuli. This makes the use of the instrument impractical or ethically questionable for use with certain populations. Fortunately, the Affinity 2.0 avoids this dilemma through utilizing clothed models in non-suggestive settings. Considering that Brown (1979) found women to rate less pornographic slides in which models were partially clothed as more arousing than more explicit pictures, the Affinity 2.0 has potential to be an especially accurate measure of sexual interest for women.

In conclusion, the Affinity 2.0, with its readily accessible data that can be norm-referenced and potential for use with a variety of populations, has emerged as a promising instrument for effectively assessing female sexual interest for both research and clinical purposes.

Statement of Problem

Given that Affinity 2.0 is a newly standardized instrument, we do not yet know what the normal heterosexual female response is to this instrument nor whether that response is temporally stable.

Statement of Purpose

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

CHAPTER III

METHOD

Participants

Criteria for inclusion. Participants in this study consisted of females with a minimum age of 18. Given that the purpose of this study is to examine VT response of normal heterosexual females, the researchers screened out all females whose sexual interest was not exclusively heterosexual as well as those who had any history of pedophilia. Homosexual interest and history of pedophilia were screened for by responses to a questionnaire that was administered to each participant following the second (re-test) administration of the Affinity 2.0. Participants who expressed homosexual interest or a history of pedophilia were as fully compensated as other participants, but were not included in the experimental group.

Location of recruitment. Participants selected for this study were sampled from the large population of students taking undergraduate psychology classes at Brigham Young University. Individuals attending psychology classes at Brigham Young University were asked to participate via a short presentation by a researcher. The researcher explained that the purpose of this study was to test a new device that claims to measure sexual interest. Potential participants were informed that participation in the study would include looking at several still images of fully-clothed models depicted in every-day life situations and rating those images on their sexual attractiveness or unattractiveness; participants repeated the process a second time two to four weeks later and then completed a short questionnaire. Compensation for any female's full participation included two free movie tickets. Participants from this pool were expected to provide a sufficiently diverse and robust sample for the purposes of this study.

Number of participants required. In past studies of sexual interest the sample sizes have not always been sufficiently robust to make desired statistical inferences. According to the Central Limit Theorem (Howell, 2002), a curve approaches normality with a sample size of $n = 30$. The larger the n, the more it will approximate the normal curve and represent the population at large. In view of these considerations, researchers decided to use a sample size of 120 individuals. A sample of this size is robust enough to make statistical inferences about the population at large, but not so great as to make the collection process unnecessarily time-consuming and expensive. Additionally, the test was administered to the same 120 participants a second time to in an effort to assess the temporal stability of the instrument.

Materials

Informed consent. Every potential participant was asked to sign an informed consent document (Appendix A). This document provided a description of the study, an assertion that the participant was void of any history of pedophilia, disclosure of what the participant would be asked to do in the study, information with regards to confidentiality, and contact information. No individual was allowed to participate in the study without first signing the informed consent document.

Sexual interest assessment. The assessment that was utilized for this study was the Affinity 2.0, a computer program developed to measure sexual interest (Glasgow, 2003). The original version of the Affinity was designed to assess the sexual interest of males with a mild learning disabilities, the current version is licensed for use as a clinical assessment tool for learning disabled adult male offenders and non-learning male offenders. Glasgow (2003) has also approved the use of 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 ten main parts (Glasgow, 2003):

1. The main screen is where the professional can identify him/herself in order to gain access to the instrument.
2. The stimulus management screen permits the user to choose which items will be used as practice items, as well as the order of those items.
3. The ‘clicker’ screen evaluates the basic mouse-pointer skills of the individual to ensure the individual’s motor skills are sufficient (as poor motor skills are likely to thwart viewing time measures).
4. The new assessment screen is where the fundamental information regarding a proposed assessment is entered (which creates a permanent record for the assessment).
5. The ranking screen provides the participant with a series of simple line drawings from which he/she is able to rank the figures as more or less attractive to him/her (through pointing and clicking on the figures with the mouse). From this data the computer is able to automatically develop a rank order of the individual’s expressed sexual preference.
6. The rating screen is where images of real males and females appear that correspond with the categories represented by the line drawings in part five (adult, adolescent, preadolescent, and small child). Participants are asked to

rate each image according to how sexually attractive/unattractive it appears to him/her. As participants view and rate each image two separate measures of viewing time are recorded.

7. The results screen presents the raw data results, providing the option to view data individually in graphical or tabular format, if desired.
8. The raw data chart screen allows the assessment results to be viewed in the form of a bar chart or table. Exploration of important features or anomalies in the bar chart view is possible as clicking on any bar in the chart will display the corresponding image.
9. The mean ranks screen shows the results on a shared axes (which have been converted to ordinal data).
10. The data management screen allows for further statistical analysis by exporting data from any number of assessments.

The participant begins the assessment by viewing and ranking several prototype images that are presented during Step 5 of the Affinity (Appendix B). These prototype images are simple line drawings that depict a character from each of the eight categories: male adult, female adult, male adolescent, female adolescent, male preadolescent, female adolescent, male small child, and female small child. The participant begins by ranking the line drawings according to their level of sexual attractiveness. When the participant reaches a point where the remaining line figures are no longer sexually attractive to her, she 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 (Glasgow, 2003).

The rating procedure consists of showing the participant a total of 56 test images and several practice images (Appendix B). 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 that ranges from "attractive" to "unattractive." As the participant is undertaking this rating procedure, two measures of viewing time are being surreptitiously recorded. The first measure of viewing time is *Viewing Time on Task* (OTL), which is the time of the 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 the time period from when the participant rates the image to the time the image is changed (Glasgow, 2003). The researcher will then calculate a third measure of viewing time, the Total Task Latency (TTL) by totaling the two viewing times (OTL+PTL). All viewing time measurements are reported in raw score form in milliseconds.

One of the major advantages of the Affinity 2.0 is that it offers to report all scores in their raw data form rather than converting the scores into ipsative z-scores, as with the AASI (Fischer & Smith, 1999). In order to truly determine how heterosexual females respond to visual stimuli, it was essential that we have access to all the raw scores for each individual. Access to this raw data allowed the researchers the opportunity to calculate category means and standard deviations for each participant which, consequently, allowed for inter-individual comparisons to be made.

Demographics, Social Desirability, and Sexual Interest Questionnaire. Participants were given a brief questionnaire called the Demographics, Social Desirability, and Sexual Interest Questionnaire (DDSQ) that was specifically designed for the purposes of this study (Appendix C). The questionnaire consisted of three sections. The first section dealt with general demographics (age, ethnicity, year in school, marital status). The second section involved a social desirability scale called the M-C 2(10) developed by Strahn and Gerbasi (1972), which is a condensed version of the Marlowe-Crowne Social Desirability Scale (M-C 33). Strahn and Gerbasi 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 the M-C 2(10) to be less offensive and more clearly worded (Mandell, n.d.). Strahan and Gebrasi (1972) found the mean score on the M-C 2(10) to be a 4.6 with a standard deviation of 2.1; scores on the M-C 2(10) above 8 indicate an extreme need for approval and concern with social desirability. The purpose of using a social desirability scale like this was to determine whether or not social desirability factors into the participants' viewing habits.

The third section determined the participant's reported sexual orientation. The sexual preference inventory included in the DDSQ was an adaptation of the Kinsey Heterosexual-Homosexual Scale (Kinsey et al., 1998) that answered whether the participant's sexual orientation was exclusively heterosexual or not (e.g., predominantly heterosexual with incidental homosexual, predominantly heterosexual with more than incidentally homosexual, equally heterosexual and homosexual, etc.).

Procedure

Confidentiality. Because sexual interest is a sensitive topic, the administration of the Affinity 2.0 and the questionnaire could be considered somewhat intrusive. Consequently, we anticipated that some students may be hesitant to respond honestly in this study unless measures were taken to ensure their confidentiality. Brigham Young University is a private religious institution and one that necessitates strict adherence to an honor code. This honor code, which all students must sign if they are to attend, prohibits behaviors such as homosexual activity and pedophilia. Failure to follow the principles of the Honor Code can result in university sanctions, possible dismissal, and even criminal prosecution (BYU, n.d.). Given that adherence to this honor code was so highly valued, we were concerned that students with homosexual feelings or pedophilic tendencies might have been reluctant to be truthful on the Affinity 2.0 and/or the questionnaire.

The informed consent document each potential individual read and signed informed the participant of the purpose, as well as expectations, for the study. The researchers also included a section that discussed the confidentiality for potential participants. It explained that a master list would be created wherein numbers would be assigned to all participant names. Upon completion of the test, re-test and questionnaire, the names of the participant were deleted. There was only one copy of this master list, which was kept in a secure, locked file cabinet. It assured potential participants that all information pertaining to the participant was kept confidential and that no names were used in the study or reported to the Honor Code Office of Brigham Young University. This informed consent document was designed to protect the individuals that participated in the study, to lessen inhibitions about being a participant, and increase the probability of honest responses on both the questionnaire and the Affinity 2.0.

Setting. Another factor that influences participant viewing time is the setting in which the Affinity 2.0. Martin (1964) discovered that participants who were asked to rate sexually explicit photos spent considerable less time looking at those photos in the presence of other people than they did when they were alone. In the Brown et al. (1973) study, participants were asked to rate a series of 15 slides that varied in their sexually explicit content. There were two different settings utilized in the experiment; one in which participants viewed the slides alone and the other in which the participant was in a room with three graduate students that reported being interested in the study. Brown et al. also discovered that participants spent a significantly shorter amount of time viewing sexually explicit slides in the presence of other people than when they viewed these slides alone.

While the rating task of the participants in this study differed from that of the Martin (1964) and Brown et al. (1973) studies in that the images in the Affinity 2.0 consist of clothed models in non-explicit poses rather than being sexually explicit stimuli, participants in this study were aware that the test they are taking was an assessment of their sexual interest. As this awareness may have led the participant to reduce viewing time in the presence of others, we decided it best that participants were alone as they completed the Affinity 2.0.

Procedure. After reading and signing the informed consent documentation, the participant was led to a private room equipped with a computer on which the Affinity 2.0 program was installed. After the participant was instructed in how to begin the program (through completing the primary prototype procedure and rating of test images), the researcher left the room so that the participant would not be influenced as she completed the rest of the assessment. The researcher informed the participant that he/she would wait outside to ensure that no one disturbed her. This setting assured the participant's privacy, which was anticipated to have lessened inhibitions and encouraged honest reactions as well as more accurate viewing times.

Upon completion of the Affinity 2.0, the participant exited the room and the researcher answered any additional questions and scheduled a second visit for two to four weeks from the time of the initial assessment. The researcher then entered the room, ensured that the date has been recorded, and prepared the room for the next participant.

When the participant returned for the re-test two to four weeks later she followed the same procedure with the Affinity 2.0 and then completed the DDSQ at the end. Upon completion of the DDSQ the participant was given two single admittance movie tickets to a local theatre as a reimbursement for her participation. Every participant in this study followed this exact procedure.

Data Analysis

The data yielded from participants during the Affinity 2.0 test and re-test can be envisioned as a three-dimensional cube that consists of 120 participants multiplied by 8 attributes multiplied by 2 administrations of the test. Hence, in total, this study yielded 1920 pieces of data for statistical analysis (120 X 8 X 2 = 1920). As we were uncertain exactly what data would result from the Affinity 2.0, it became necessary to employ a variety of data analyses in ascertaining which analyses will best described/fit the data. For the results of our analyses we will employ the conventional predetermined significance level of $p < .05$, those results that surpass this threshold will be considered significant (Howell, 2002).

Sums. We obtained totals for each attribute/category by summing the total time an individual spent viewing all seven slides in a category. We anticipated obtaining three different sums: OTL (On-Task Latency), PTL (Post-Task Latency), and TTL (Total Task Latency). The Affinity 2.0 calculates both the OTL and PTL. In addition, we proposed to calculate the TTL by adding the OTL and PTL viewing times together. The inferential statistic that will be utilized for comparing the category sums from test to retest is the Pearson Product Moment Correlation Coefficient (PPMCC).

Means. After obtaining the sums for OTL, PTL and TTL we then obtained the mean time spent viewing slides in each category by dividing the sum of each category by 7 to yield the average time spent viewing a slide in that category. The mean was expected to be an effective way to describe the central tendency of our distribution. Again, the appropriate inferential statistic is the PPMCC.

Medians. We also determined the median viewing time for each category as it was conceivable that viewing time could be altered if a participant dropped the mouse, became distracted, sneezed, etc. In these cases, where the data may become skewed, we determined that it would be advantageous to calculate the median as it may be a better measure of central tendency. The appropriate inferential statistic for comparing the medians obtained at test and retest is the Spearman's Rho Correlation.

Ipsative scores. Whereas the first three methods were independent in nature, the next four methods of intended statistical analysis all utilized ipsative scores. Since the use of ipsatives is not as widespread as other methods of data analysis, a brief discussion of the nature of ipsative scores seems warranted.

Ipsative scores involve a scale conversion of data that allows one to compare differences between attributes intra-individually. To obtain ipsative scores, (1) sum the raw scores across attributes, (2) choose an arbitrary constant to sum to, (3) subtract the individual total from the constant, (4) divide the result by the number of attributes, and (5) subtract the resulting quotient from each raw score. The resulting ipsative scores all sum to the constant and, thus, are necessarily dependent upon one another (Clemans, 1956).

A simple way to think about ipsative scores is to imagine you have a limited (constant) number of chips that you are free to distribute across a certain number of attributes. If you “spend” 35 of your chips on the first attribute then that leaves only 65 chips to divide among the remaining categories. If you only “spend” 2 of your chips on another category it stands to reason that you likely preferred the former attribute to the latter. This method is helpful in that it can reveal the differences between attributes within an individual. This technique is limited in that it transforms results into ordinal data and also precludes inter-individual comparisons.

Category ranks. One ipsative technique we anticipated would be helpful in analyzing our data was that of summing the raw scores across categories and ranking each on a scale of 1-8 according to viewing time. The ipsative constant for ranking the different categories was 36 (8 + 7 + 6 + 5 + 4 + 3 + 2 + 1 = 36). This ranking transformed the results into ordinal data that yielded an idea of relative position for the slides in each category (rather than the absolute scores). The appropriate inferential statistic was Spearman's Rho Correlation.

Affinity 2.0 weighted ranks. In addition to knowing which categories a participant viewed longest, we were also curious which of the individual slides were viewed for the longest amount of time. Fortunately, the Affinity 2.0 performed this operation for us by creating weighted ranks for each of the slides. This was accomplished by assigning a rank to the raw OTL of each slide, 1-56 (with 56 being assigned to the longest viewed slide and 1 being assigned to the shortest viewed image). The mean ranks are obtained by averaging the ranks assigned to each slide in any given category. The lowest possible mean rank assigned to any given category is 4 [(1 + 2 + 3 + 4 + 5 + 6 + 7) / 7 = 4]. The highest possible mean rank that can be assigned to any category is 53 [(56 + 55 + 54 + 53 + 52 + 51 + 50) / 7 = 53]. The sum of the mean ranks of the eight categories is a constant 228. After obtaining the mean ranks, the values at time one were correlated with the values at time two using Spearman's Rho Correlation.

Standardized ipsative scores. An aforementioned drawback to ipsative scores is that they preclude inter-individual comparisons. However, Clemans (1956) claimed that there is a solution to this dilemma. He proposed three steps for making ipsative scores commensurable so that one might make inter-individual comparisons: (1) convert raw scores for each attribute into t-scores, (2) transform the resulting scores into ipsatives scores, and then (3) standardize the ipsatized scores.

While ideal in theory we found that, after running tests with "dummy data," this technique appeared to distort the data. While the usefulness of this technique was questionable for our purposes in describing our sample's response to the Affinity 2.0, due to the resulting distortion of scores, we still determined to conduct this analysis to see if it yielded any useful results.

Standardized ipsatized OTL raw scores. Another approach to standardizing ipsative scores is that of ipsatizing and standardizing OTL raw scores. This analysis is identical to that described above with the exception of omitting the first step in which raw scores are transformed into t-scores. In running this analysis with “dummy data” (using a constant of 120) it was discovered that this technique more fully maintained the integrity of the data; thus, we decided to see if this analysis might prove useful for our purposes.

Chi-square analysis for temporal stability. Since the Affinity 2.0 is a relatively new instrument we wanted to assess its temporal stability. Given that we assume viewing time to be a stable behavior that is not likely to change significantly over brief periods of time, we expected for the re-test to be similar to the initial test results for each individual. Hence we conducted a “goodness of fit” chi-square statistical analysis wherein the mean ipsative weighted rank value for each category (calculated from the seven slides in each category) obtained during the initial assessment served as the expected (*E*), and the re-test results served as the observed (*O*) for:

$$\chi^2 = \frac{\lfloor O\text{-}E \rfloor^2}{E} \ .$$

If the results of this analysis reveal that the results obtained from the second viewing (*O*) are not significantly different from the expected (*E*), we will safely conclude that the heterosexual female response to the Affinity 2.0 was a temporally stable one.

In conclusion, through utilizing various parametric and non-parametric statistics, we wished to discover which of these analyses best described/fit the data yielded by our sample to the Affinity 2.0. In addition, through using a “goodness of fit” chi-square to analyze the test and re-test results, we were able to discover whether our sample’s response to the Affinity 2.0 was a temporally stable one.

CHAPTER IV

RESULTS

Participant Demographics

There were 146 female participants who took the Affinity 2.0 at time one (test). One hundred twenty of these original 146 participants returned to retest, answered the items on the DDSQ, and fulfilled all inclusion requirements for this study. The mean age of these 120 participants was 21.67 years (*SD* = 5.25; minimum 18 years, maximum 56 years). One hundred and four of the participants were Caucasian/White, 4 were Hispanic, 3 were Asian American, 2 Multiracial, 2 Native American, and 1 from each of the following nationalities: European-Salvadorian, Romanian, White Peruvian, Scandinavian, and South African. Eighteen (15%) of participants were currently in their freshman year of college, 24 (20%) were sophomores, 28 (23.3%) were juniors, 41 (34.2%) were seniors, and 9 (7.5%) were graduate students. Eighty-nine of the participants were single (74.2%), 29 were married (24.2%), and 2 (1.7%) were divorced; none were widowed. Participants scored in the average range (mean = 4.02; *SD* = 1.9) on the Marlowe-Crowne Social Desirability Scale-10 indicating that our sample was not overly concerned with social desirability (Strahan & Gerbasi, 1972). All 120 participants met the inclusion criteria of marking their sexual interest as “Exclusively heterosexual with no homosexual” on the Kinsey Scale.

Sums

We proposed to conduct three separate analyses of our sample’s response to the Affinity 2.0 based on OTL, PTL, and TTL. It was hypothesized that PTL and TTL would provide additional, unique information concerning participant’s VT that, in combination with an analysis of OTL, would yield a more complete picture of participants' behavior and sexual interest. However, an analysis of PTL and TTL did not support this assumption.

The average range of PTL across images/categories was .76 - .91 seconds—an average difference of only .15 seconds. It appears that after the rating task was completed (measured by OTL), participants moved quickly onto the next slide rather than spending additional time pausing or lingering on an image (measured by PTL). Such a small and consistent amount of time did not prove to offer any additional, useful information regarding participant behavior (VT) and/or sexual interest as PTL was virtually identical to OTL. Given the miniscule and consistent nature of PTL, the ensuing sum of TTL emerged as a mere surrogate for OTL, yielding no further meaningful data. This result is almost identical to Sneed (2006) where analyses of PTL and TTL were also discarded due to their insignificant nature.

The purpose of this study was to determine and describe a typical heterosexual female response to Affinity 2.0; given that any analyses conducted with PTL and TTL fail to assist us in this endeavor given their repetitive and futile nature, all analyses with PTL and TTL were abandoned, including a comparison of sums. All subsequent results reported are based solely on analysis of OTL.

Means

Based on OTL, we computed the means, standard deviation, and ranges for our sample's response to the Affinity 2.0 at both test and retest (see Table 1). A comparison reveals a decrease in all mean times and standard deviations from test to retest as well as the presence of a consistent preference in VT. At both test and retest ADM (adult male) images were the most preferred visual stimuli followed by images of JUM (juvenile male). ADF (adult female) and JUF (juvenile female) images, respectively, were the next longest viewed images in both trials followed in turn by PJM (pre-juvenile male) images. PJF (pre-juvenile female) and SCF (small child female) images were viewed for almost identical amounts of time at both test and retest, while SCM (small child male) images were viewed for the least amount of time in both trials.

Table 1

Affinity 2.0 Mean OTL (sec)—Test & Retest

Category	Test (*n* = 120)			Retest (*n* = 120)		
	Mean	*SD*	Range	Mean	*SD*	Range
ADM	4.23	1.76	13.67	3.26	1.35	9.35
JUM	3.51	1.71	17.54	2.68	0.91	5.46
PJM	2.35	0.95	6.73	1.84	0.81	7.49
SCM	2.00	0.95	7.57	1.61	0.47	3.63
ADF	2.72	1.24	7.73	2.24	1.11	7.65
JUF	2.49	1.11	8.07	2.00	0.90	6.83
PJF	2.09	0.90	6.33	1.69	0.76	7.75
SCF	2.10	0.73	4.50	1.70	0.45	3.23

Figure 1 provides a visual depiction of the Female Mean Raw Score OTL Response to Affinity 2.0 at test. Given that the identical ordering of the category VT preferences at both test and retest, this curve is representative of our normal heterosexual female response to the Affinity 2.0 at both test and retest. As can be observed, the VT standard deviations surrounding the adult and juvenile categories for both sexes are larger than those surrounding the pre-juvenile or small child categories. This occurrence was evident in both trials and seems to attest to the normalcy of our sample (i.e, the reduction of VT for pre-juvenile and small child images suggests a lack of sexual interest in children).

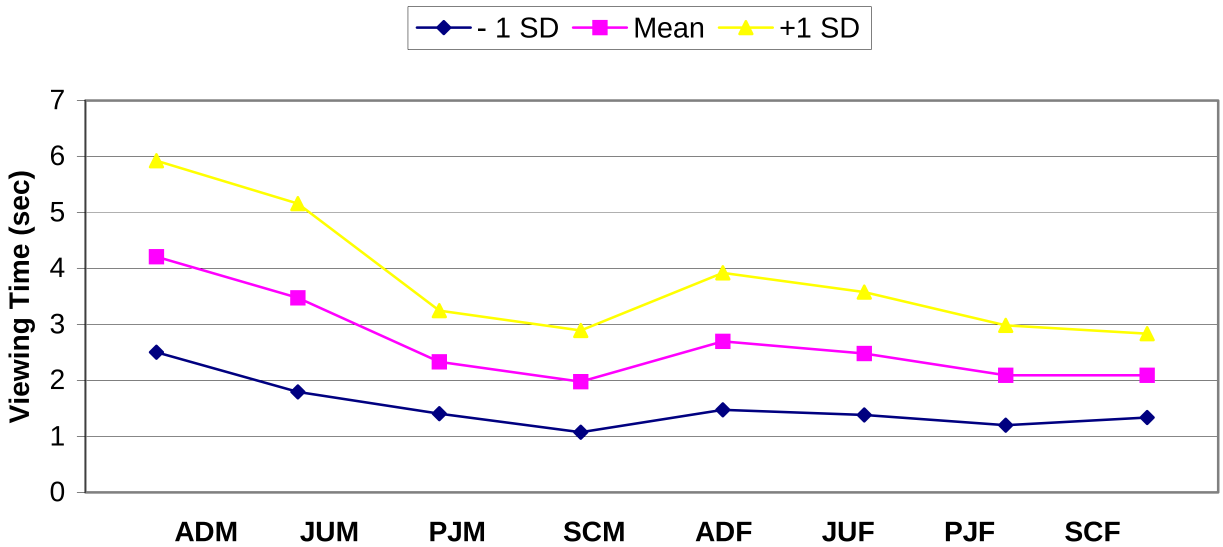

Figure 1. Female mean OTL raw score response curve to Affinity 2.0—test

Figure 2 juxtapositions the mean OTL raw response curves from test to retest. The decrease in VT overall from time one to time two is evident as well as the consistent ordering of VT category preferences. Both test and retest suggest the existence of a discernable VT response curve for our sample of normal heterosexual females.

After obtaining the OTL means, a Pearson Product Moment Correlation coefficient was calculated to determine the relationship between participants' OTL at both test and retest. The results of this analysis are represented below in Table 2. As can be seen, while all coefficients proved to be statistically significant at the $p < .01$ level, they can be considered only mildly to moderately strong in their practical significance. These weak to moderate correlations suggest that these women were not very consistent in their VT behavior of Affinity images from test to retest.

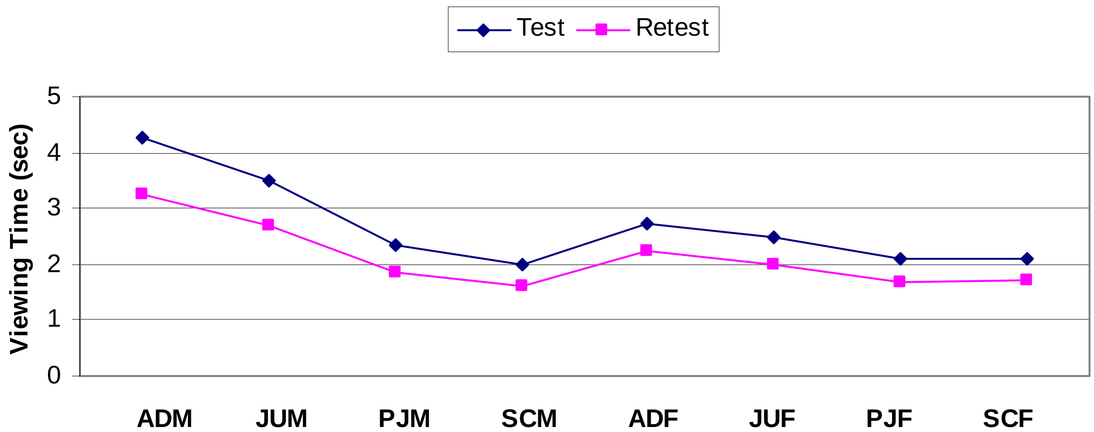

Figure 2. Female mean OTL raw score response curves to Affinity 2.0—Test and retest

Table 2

Pearson Correlation of Affinity 2.0 Mean OTL—Test to Retest

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
r	0.63**	0.26**	0.28**	0.36**	0.38**	0.30**	0 .52**	0.37**

Note. **. Correlation is significant at the 0.01 level (2-tailed). *N* = 120

Medians

Our next analysis consisted of obtaining medians for OTL VT to help us determine if medians were a better measure for describing the central tendency of our sample's response to the Affinity 2.0, (i.e., to help correct for any possible skew in our data). Given that, when compared to the means, our medians are lower in value, it attests to the fact that our data were positively skewed overall. Considering the rating task participants are asked to perform and the fact we would expect participants to look at some slides longer than others (due to an interest in a particular image), this skew is predictable and expected. Since it is likely that responses to the Affinity 2.0 will contain this positive skew, it seems advisable that medians be referenced as they provide a more accurate measure of central tendency for skewed data than do means. The average medians from test and retest are shown in Table 3.

Table 3

Mean Median—Affinity 2.0—Test to Retest

Category	Test (*n* = 120)	Retest (*n* = 120)
	Mean	Mean
ADM	3.91	2.91
JUM	2.90	2.30
PJM	2.08	1.64
SCM	1.77	1.49
ADF	2.42	2.11
JUF	2.23	1.75
PJF	1.87	1.52
SCF	1.85	1.53

As can be observed, like the means, these scores also reflect an overall drop in the median VT from time one to time two. In addition, a consistent preference in VT also is apparent. In both trials ADM images were the most preferred visual stimuli with JUM images were the second longest viewed in both trials followed by ADF and JUF images. PJM images, PJF images and SCF images, respectively, were the next longest viewed with SCM images being viewed for the least amount of time at both test and retest. This ordering of preferred VT categories is identical to the ordering found with the means.

Figure 3 provides a visual representation of the overall average median OTL VTs for our sample at both test and retest. Despite the decrease in median VT from test to retest, the overall VT response curve for both trials is very similar as the ordering of the preferred visual stimuli categories is identical.

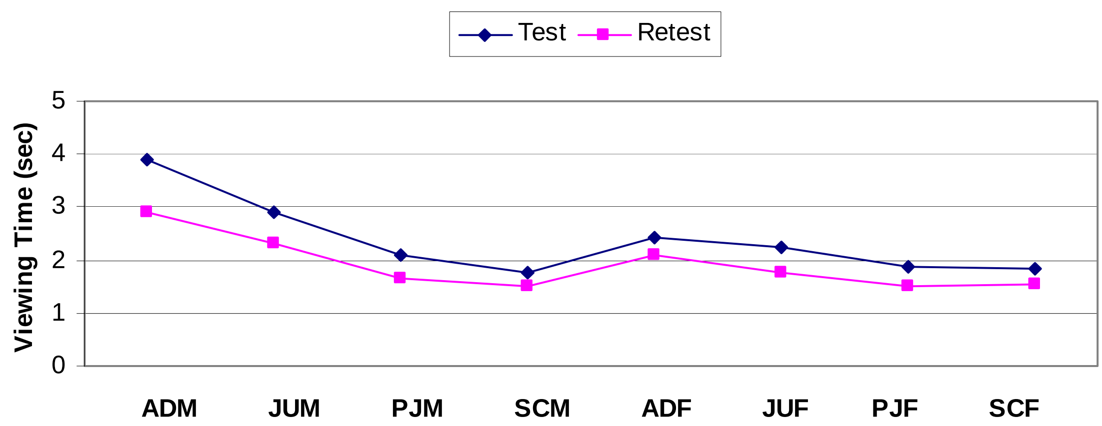

Figure 3. Average median response to Affinity 2.0—Test and retest

After calculating medians for participants we ran a Spearman's Rho correlation between medians at test and retest. The results of this analysis can be observed below in Table 4. With the exception of two categories (ADM and PJF), the Spearman's Rho correlations proved to be stronger that the Pearson's correlations. Hence, we can confirm that medians are a more consistent measure of central tendency for our data and demonstrate more stability than the mean OTL. Like the Pearson correlations, we found all of the

Spearman's Rho correlations to be significant at the p < .01 level; unlike the Pearson correlations, however, the Spearman's Rho correlations begin to approach practical significance.

Table 4

Spearman's Rho Correlation of Affinity 2.0 Median OTL—Test to Retest

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
r	0.46**	0.53**	0.47**	0.43**	0.65**	0.58**	0.41**	0.41**

Note. **. Correlation is significant at the 0.01 level (2-tailed). *N* = 120

Ipsative scores

Category ranks. The first ipsative score analyzed was that of category ranks. For this analysis we first summed the total amount of time each participant spent looking at each image of the seven images in a category. After obtaining the sums for each of the eight categories we then ranked the categories 1-8 according to the amount of VT spent communally on each. A "1" represented the category with the highest VT, a "2" denoted the category viewed for the second longest period of total time, "3" represented the third longest collective VT, and so on, with "8" representing the category with the least amount of VT. Any ties in total VT were not broken and were recorded as such. After assigning these ranks, a Spearman's Rho correlation was between the category ranks for each individual from time one to time two. The results of this analysis are recorded in Table 5 below.

As can be seen, only two categories, JUM and ADF, proved to be statistically significant out of all the categories. And while the correlations for JUM and ADF proved to be statistically significant at the 0.01 level, they would still be considered quite weak. Correlations for the other categories were found to be extremely weak—especially as two of the categories, JUF and SCF, were actually negatively correlated. It seems as if the technique of transforming the raw OTL VT times into category ranks was too crude and created too much distortion to give a helpful description of our data.

Table 5

Spearman's Rho Correlation of Affinity 2.0 Category Ranks—Test to Retest

Category	ADMJUM	PJM	SCM	ADF	JUF	PJF	SCF	
r	0.14	0.26**	0.01	0.09	0.31**	-0.04	0.01	-0.02

Note. **. Correlation is significant at the 0.01 level (2-tailed). *N* = 120

Affinity 2.0 ipsative weighted ranks. Our next analysis concerned the Affinity 2.0 ipsative weighted ranks. We calculated the mean rank for each category as well as the standard deviation and range for both test and retest. The results of these analyses can be seen in Table 6.

As is illustrated in Table 6, a consistency from time one to time two in terms of weighted ranks was found. An identical ordering of categories was found with ADM being ranked highest at both trials followed in turn by JUM, ADF, JUF, PJM, SCF, PJF, and SCM, respectively. The standard deviations and ranges were also similar from time one to time two (with the largest discrepancy in standard deviations from test to retest being .82 in the ADM category and the largest difference in range being 4.58 in the PJF).

Table 6

Affinity 2.0 Weighted Ranks

Category	Test (*n* = 120)			Retest (*n* = 117)		
	Mean	*SD*	Range	Mean	*SD*	Range
ADM	43.60	6.74	28.28	43.37	5.92	29.14
JUM	36.77	6.28	29.21	35.91	6.16	33.00
PJM	26.44	5.45	26.14	24.76	5.52	24.50
SCM	19.69	5.62	28.28	20.72	5.74	26.15
ADF	29.95	7.86	36.29	31.73	7.84	37.00
JUF	27.82	6.20	31.86	27.68	6.28	32.79
PJF	21.51	6.09	30.29	21.61	5.43	25.71
SCF	22.13	5.09	24.93	22.21	5.85	26.28

Figure 4 is a visual representation of the normal heterosexual female ipsative weighted ranks response to the Affinity 2.0. This curve shows the mean ipsative weighted rank response as well as one standard deviation above and below.

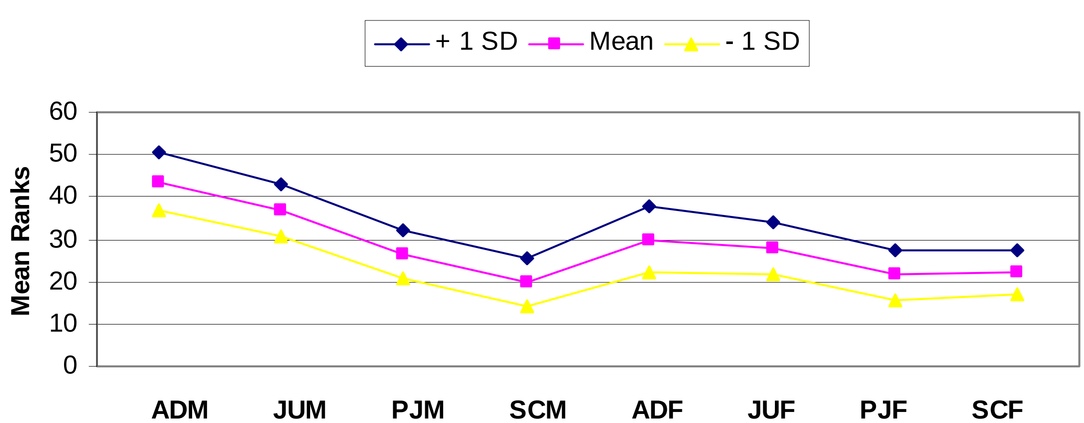

Figure 4. Female mean ipsative weighted rank response curve to Affinity 2.0—Test

It is easy to note the similarity between the ipsative weighted rank response curve to that of the mean OTL raw score response curve. It appears that the Affinity ipsative weighted ranks seem to create less distortion of actual VT behavior as the integrity of the OTL raw score response curve was maintained as well as the ordering of preferred VT categories. Unlike category ranks, it appears that this ipsative measure is a helpful way to describe our sample's response to the Affinity 2.0

Figure 5 juxtaposes the mean ranks responses from test and retest. Because the Affinity 2.0 weighted ranks are ipsative measures based on intra-individual response rather than on actual behaviors (like OTL mean and medians), the response curves for test and retest are practically indistinguishable with preferred category orderings that are identical from time one to time two.

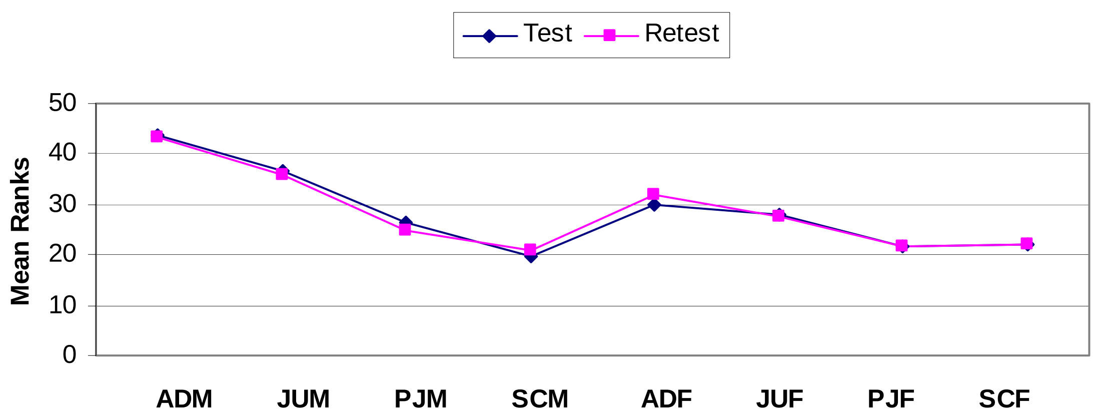

Figure 5. Female mean ipsative weighted rank response curve to Affinity 2.0—Test to retest

We next calculated Spearman's Rho correlation for the ipsatized weighted mean ranks. The results of this analysis are below in Table 7. As is shown, the correlations for

ADM, JUM, and ADF images were all statistically significant at the 0.01 level. The correlation for JUF was significant at the 0.05 level. However, none of these statistically significant correlations are above moderate in their strength. Of the correlations for the rest of the categories, all are very weak with PJF being negatively correlated.

Table 7

Spearman's Rho Correlation of Affinity 2.0 Ipsative Weighted Ranks—Test to Retest

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
r	0.44**	0.29**	0.06	0.18	0.41**	0.20*	-0.03	0.16

Note. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed). *N* = 117

Thus, overall, ADM and ADF images were the most consistently ranked categories by the women in our sample followed by the JUM and JUF categories. The ranking of the pre-juvenile and small child categories by women in our sample proved to be significantly less consistent.

Standardized ipsatized scores. In attempting to utilize standardized ipsatized scores we affirmed, just as with our experimentations with dummy data, that this technique created too much distortion (as the standardized ipsatized scores failed to even maintain the ordering of categories found in the actual VT behavior of our sample). Given this distortion, we determined that any further analyses of standardized ipsatized scores would not be of any use to us in our attempt to determine and describe a normal heterosexual female response to Affinity 2.0. Thus, we discarded any attempts to make use of standardized ipsatized scores.

Standardized ipsatized OTL raw scores. The last ipsative score analysis we planned to utilize in determining and describing our sample's response to the Affinity 2.0 was that of standardized ipsatized OTL raw scores. As this technique does not require the transformation of raw data into t-scores we hoped it would more fully maintain the integrity of the data. However, in our attempt to employ this technique we discovered that it, just like the standardized ipsatized scores, also resulted in too much distortion. Thus, the use of standardized ipsatized OTL raw scores was also abandoned as they did not prove helpful for our purposes.

Chi-Square for Temporal Stability

We proposed to utilize a Chi-Square as an estimate of temporal stability using Test for the expected (*E*) and Retest for the observed (*O*) values. Given the dependent nature of these scores (that they add up to a constant), a chi-square goodness of fit seems a viable option for determining temporal stability. If the overall response to the Affinity 2.0 is temporally stable then we expect to obtain a chi-square value of under 14 (7 *df*) with no residuals exceeding $\pm$ 1.96.

A chi-square goodness of fit test was calculated comparing the Affinity 2.0 weighted ranks from retest to those obtained at test. The results of this analysis are shown in Table 8 below.

It was hypothesized that (if sexual interest is a temporally stable construct) the results at retest would not differ significantly from those obtained at test. No significant deviation from the critical value was found ($\chi^2(7) = 0.29$, $p < .05$). Consequently, it can be observed that none of the residuals for any of the eight categories surpassed the significant $\pm$ 1.96 level, indicating that none of the residuals contributed significantly to our chi-square value. Data from retest seem consistent with the expected values obtained at test; thus, we can safely conclude that the overall average normal heterosexual female response was temporally stable from test to retest.

Table 8

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

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF	TOTAL
Observed	43.37	35.91	24.76	20.72	31.73	27.68	21.61	22.21	227.99
Expected	43.60	36.77	26.44	19.69	29.95	27.82	21.51	22.13	227.91
O-E	-0.23	-0.86	-1.68	1.03	1.78	-0.14	0.10	0.08	0.08
O-E SQ	0.05	0.74	2.82	1.06	3.17	0.02	0.01	0.006	
O-E SQ/E	0.001	0.02	0.11	0.05	0.11	0.0007	0.0005	0.0003	0.29
Residual	-0.03	-0.14	-0.33	0.23	0.33	-0.03	0.02	0.02	

Utilizing this chi-square goodness of fit procedure could potentially be a valuable way of ascertaining overall temporal stability in the future whether on a group, individual, case-by-case, or comparison basis. Further, this analysis can help ascertain whether an individual's overall responses to Affinity 2.0 from test to retest are stable and could thus be a means to potentially detect dissimulation.

Additional Analyses

In addition to utilizing the techniques for determining and describing our sample's response to the Affinity 2.0 initially proposed, we determined to conduct additional analyses that included the participants' self-report. We wished to see how the participants' self-report (ratings) related to an ordinal measure of their actual behavior (VT), which prompted an examination of our obtained ratings and the stability of those self-report ratings from test to retest. Since we did not know beforehand whether raw or ipsatized measures would provide the most helpful description of our data, we followed the same pattern we did in the proposed analyses by analyzing the raw data first prior to analysis of the ipsatized scores.

Raw mean rating and raw mean rank OTL. Our first additional inquiry was to see how self-report related to VT behavior. Table 9 provides an analysis of the correlation between raw mean rating and raw mean OTL for our sample at test and retest.

Table 9

Spearman's Rho Correlation of Raw Mean Rating and Raw Mean Rank OTL

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
				Test (*n* = 120)				
r	-.19*	0.19*	0.49**	0.43**	0.56**	0.58**	0.41**	0.40**
				Retest (*n* = 120)				
r	-0.12	0.46*	0.27*	0.09	0.55**	0.28**	0.07	0.09

Note. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

As can be seen, though a majority of the correlations were significant at the $p < .01$ and $p < .05$ levels, only a handful would be considered moderately strong while most would be considered weak. While it was not surprising to find that the majority of the correlations possessed statistical, and lacked practical, significance, we did find the negatively correlated ADM category to be a bit perplexing. To help us make sense of these results, we decided to conduct an analysis of our raw ratings as well as the stability of those ratings from test to retest.

Analyses of raw rating scores. The first step in our analysis of ratings was to examine the minimum and maximum raw ratings in addition to the mean, SD and range of our sample at both test and retest. Table 10 provides a summary of this information.

As can be seen, the only category to register above a neutral response (i.e., a score of “0” on the sexually attractive rating scale) was ADM; it was clearly the highest rated category of all and was also the only category to never receive a score “-7.00” (the lowest score possible on the sexually attractive—not attractive scale). As can be observed, all other categories were negatively rated by our sample with JUM, ADF and PJF, respectively, being rated higher than any of the pre-juvenile or small child categories. Though ADM was the only category to average a positive response in ratings, the mean rating for ADM images was still relatively low on the sexual attractiveness rating scale (a 1.95 out of a potential 7).

Previous analyses have demonstrated the existence of longer VT for ADM images; however, in referencing our findings on rating scores, we discover that these longer VTs did not translate into the higher ratings that one might expect. Thus, the fact that ADM was the only category to yield a negative correlation can be attributed to the combination of considerably longer VTs and moderately low ratings on the sexual attractiveness scale.

Figure 6 is a visual representation of the data presented in the top portion of Table 9 (test) and includes the mean raw score response as well as one standard deviation above and below.

In viewing Figure 6, it is interesting to note how the raw ratings curve follows the same pattern we have seen in other previous VT curves (mean, median, ipsatized weighted ranks, etc.). True to form, we observed that ADM images were rated highest followed by JUM images. The slight elevation in ADF and JUF images, respectively, was apparent, which resulted in them being the next highest rated categories. All pre-juvenile and small child images were rated the lowest of all.

Table 10

Raw Ratings Affinity 2.0—Test to Retest

Category	Mean	SD	Range	Minimum	Maximum
			Test (n=120)		
ADM	1.95	2.07	10.86	-5.00	5.86
JUM	-2.83	2.25	10.71	-7.00	3.71
PJM	-5.75	2.28	8.29	-7.00	1.29
SCM	-5.94	2.19	7.71	-7.00	0.71
ADF	-4.34	3.36	11.29	-7.00	4.29
JUF	-5.40	2.34	7.71	-7.00	0.71
PJF	-6.03	2.00	7.57	-7.00	0.57
SCF	-6.15	1.94	8.00	-7.00	1.00
			Retest (n=120)		
Category	Mean	SD	Range	Minimum	Maximum
ADM	2.37	2.17	11.57	-5.57	6.00
JUM	-2.88	2.16	9.14	-7.00	2.14
PJM	-5.93	2.12	7.29	-7.00	0.29
SCM	-6.08	2.11	7.57	-7.00	0.57
ADF	-4.35	3.33	11.86	-7.00	4.86
JUF	-5.54	2.19	8.29	-7.00	1.29
PJF	-6.07	1.97	7.00	-7.00	0.00
SCF	-6.20	1.93	7.14	-7.00	0.14

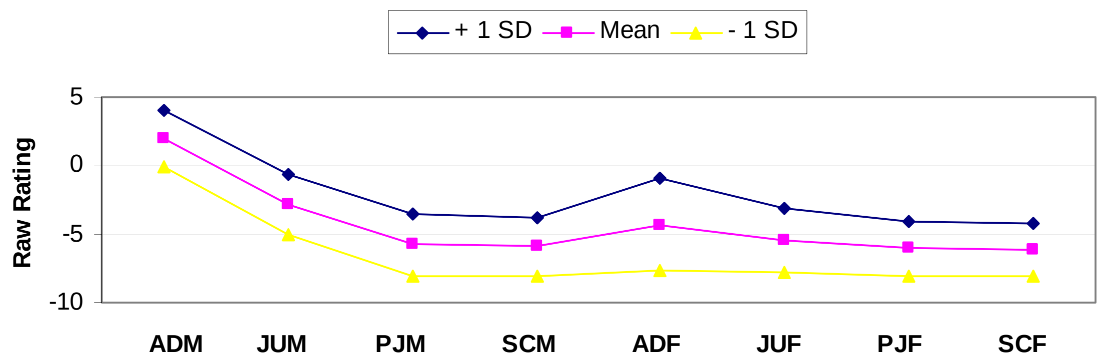

Figure 6. Raw ratings response curve to the Affinity 2.0 test

Figure 7 shows the juxtaposed raw ratings curves from test to retest. As can be observed, the curves are practically identical. Thus, the overall raw score rating responses were extremely stable from test to retest.

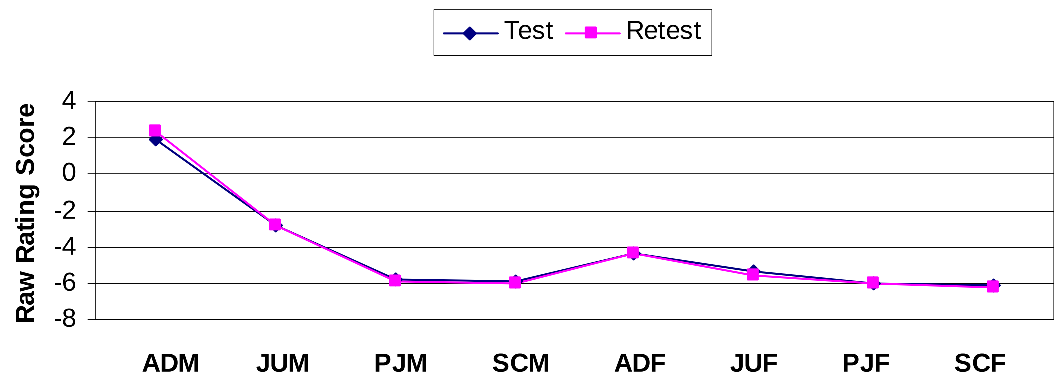

Figure 7. Raw ratings response curve to the Affinity 2.0—Test to retest

To help us ascertain exactly how stable the raw ratings were from test to retest we ran a Spearman's Rho correlation, the results of which can be seen in Table 11 below. As can be observed, the correlations for the raw ratings from test to retest were the strongest correlations obtained yet and were practically, as well as statistically, significant as most (with the exception of SCM and PJF) would be considered strongly correlated. These correlations reflect the stability in our sample in their self-report ratings of images from test to retest.

Table 11

Spearman's Rho Correlation of Affinity 2.0 Raw Ratings—Test to Retest

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
r	0.81**	0.78**	0.72**	0.64**	0.83**	0.79**	0.66**	0.77**

Note. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed). *N* = 120

As our sample consisted of women who claimed to be exclusively heterosexual, it seemed puzzling that a slight elevation would exist for ADF and JUF slides, and that some ratings for pre-juvenile and small child images were above the minimum. A qualitative piece of information that may help to provide an explanation for this phenomenon was obtained from a participant who wrote the following as an addendum to the DDSQ:

> P.S. The first time I took the assessment I didn't really understand what was meant by "sexually attractive." I should have asked, I was thinking more along the lines of "good looking" which was also "cute" for kids, rather than if I would want to sleep with them. My second round answers reflect who I think is "sexually attractive" (that I would sleep with.) Just thought you should know. Thanks!

It is unknown how many participants may have made this same error when rating the images in Affinity 2.0. Thus, rating information may be a bit unreliable as certain images might have been rated on a basis other than sexual attractiveness. Another explanation for this phenomena could be that our sample of normal heterosexual women were simply less repulsed by the possibility of having an adult or juvenile female as a potential sexual partner than they were by the notion of having a pre-juvenile or small child of either sex. If true, this would attest to the normalcy (i.e., non-pedophilic nature) of our sample. Nonetheless, as the basis for ratings is indefinite, it stands to reason that this information should be interpreted with caution.

Ipsatized mean ratings and ipsatized mean rank OTL. After looking at the raw scores, we determined to see how the ipsative ratings compared and which measure would best describe our sample's response to Affinity 2.0. Again, we were curious how the ipsatized ratings were related to the ipsatized mean ranks. Table 12 provides the results of this analysis at test and retest.

Table 12

Spearman's Rho Correlation of Ipsatized Mean Rating vs. Ipsatized Mean Rank OTL

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
				Test (n=120)				
r	0.23*	0.42**	0.28**	0.25**	0.45**	0.38**	0.11	0.12
				Retest (n=117)				
r	0.21*	0.46*	0.28**	0.04	0.59**	0.19*	0.18*	0.17

Note. ** Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

Similar to the raw ratings and OTL, the ipsatized rating scores and mean ranks also produced weak correlations that possessed little practical significance despite many being statistically significant. These results also suggest an inconsistency in our sample between ratings and VT. However, a notable consideration in interpreting this data is the quandary of potential misratings discussed above (where some individuals rated images on a basis other than sexual attractiveness to the participant). Given that it is impossible to discern the occurrence or confounding influence of such misratings, caution should be used as it is unknown how accurate or trustworthy these correlations may be.

Analyses of Ipsatized Ratings. We next analyzed the ipsatized ratings obtained from our sample's response to the Affinity 2.0. Table 13 compares the means, SD, and ranges obtained from time one to time two from our sample. The results at test and retest are practically indistinguishable with ADM being rated the highest followed by JUM, ADF, JUF, PJM, SCM, PJF, and SCF in both trials.

Figure 8 is a visual representation of the ipsatized weighted ratings from test one showing one standard deviation above and below the mean. As can be seen in viewing Figure 8, the transformation of the raw ratings into ipsative mean ratings appears to create minimal distortion from actual behavior to ipsative. It seems a good approximation of raw score ratings given that the raw rating and ipsative mean rating curves are essentially identical.

Figure 9 juxtapositions the ipsatized weighted rating response curves from test to retest. As can be observed, the curves were practically indistinguishable.

Table 13

Affinity 2.0 Ipsatized Mean Ratings—Test to Retest

Category	Test (*n* = 120)			Retest (*n* = 117)		
	Mean	*SD*	Range	Mean	*SD*	Range
ADM	49.40	5.25	31.07	49.20	5.52	28.86
JUM	35.32	6.46	36.72	34.62	6.13	31.43
PJM	23.48	4.24	25.86	23.08	4.21	23.00
SCM	21.71	4.71	31.43	21.39	4.81	26.43
ADF	31.22	9.14	36.36	32.22	9.08	33.36
JUF	25.11	4.55	23.07	25.28	3.92	23.93
PJF	21.26	3.65	19.93	21.71	3.32	18.43
SCF	20.59	4.62	21.85	20.47	4.79	22.50

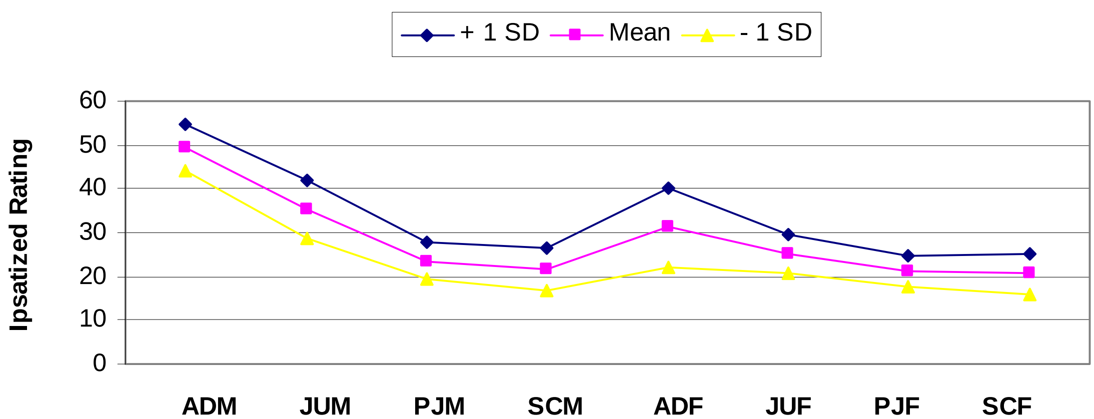

Figure 8. Ipsatized weighted ratings response curve—Test

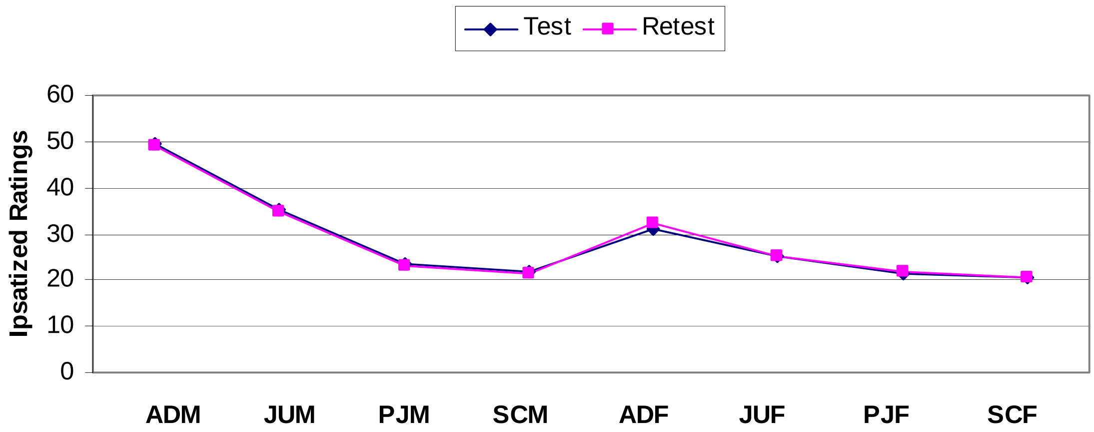

Figure 9. Ipsatized weighted ratings response curve—Test to retest

As could be inferred from Table 13, the curves from test to retest are practically indistinguishable from one another as the ratings (from highest to lowest). Both trials were ADM, JUM, ADF, JUF, PJM, SCM, PJF, and SCF.

We next wanted to determine the stability of ipsatized ratings from time one to time two. We ran a Spearman's Rho correlation to determine how stable the ipsatized mean ratings were, the results are shown below in Table 14. With the exception of one category (SCM), the ipsatized ratings had lower correlations than did their raw score rating counterparts (refer to Table 11). This reveals that the ipsatization of the raw rating scores caused a slight distortion in our data resulting in more conservative estimates of the relationship of ratings at test and retest. Despite this fact, the ipsatized mean rating correlations still demonstrated practical significance (moderately strong correlations) in addition to statistical significance. Compared to VT, it seems that the self-report of our sample of normal heterosexual females was the most stable behavior of all.

Table 14

Spearman's Rho Correlation Affinity 2.0 Ipsatized Mean Ratings—Test to Retest

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF
r	0.57**	0.67**	0.65**	0.67**	0.73**	0.53*	0.65**	0.72**

Note. **Correlation is significant at the 0.01 level (2-tailed). *N* = 117

Chi-Square Goodness of Fit Test for Temporal Stability of Ipsatized Rating Scores.

Given that our ipsatized ratings sum to a constant (like the ipsatized weighted mean ranks), we decided to run a chi-square goodness of fit test to determine the overall temporal stability of our sample's ratings from time one to time two. We used the weighted ratings obtained at test as the expected values (*E*) while utilizing the ipsatized weighted ratings obtained at retest for the observed values (*O*). The results of this analysis can be seen below in Table 15.

A chi-square goodness of fit test was calculated comparing the Affinity 2.0 ratings from retest to those obtained at test. Assuming that sexual interest is a stable construct, it was hypothesized that, if the results from Affinity 2.0 were temporally stable, that the rating results at retest would not differ significantly from those obtained at test. No significant deviation from the hypothesized values was found ($\chi^2(7) = 0.07$, $p < .05$). Data from retest seem consistent with the expected values obtained at test; thus, it can be concluded that the ratings from test to retest are temporally stable.

An even lower chi-square value was obtained for the ipsatized ratings than for the ipsatized mean ranks. This result indicates that the self-reported ratings of our sample were even more stable than their VT responses from test to retest.

Table 15

Affinity 2.0 Ratings Chi-Square Estimate of Temporal Stability—Test to Retest

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF	TOTAL
Observed	49.20	34.62	23.08	21.39	32.22	25.28	21.71	20.47	227.97
Expected	49.40	35.32	23.48	21.71	31.22	25.11	21.26	20.6	228.10
O-E	-0.20	-0.70	-0.40	-0.32	1.00	0.17	0.45	-0.13	-0.13
O-E SQ	0.04	0.49	0.16	0.10	1.00	0.03	0.20	0.02	
O-E SQ/E	0.0008	0.01	0.007	0.005	0.03	0.001	0.01	0.0008	0.07
Residual	-0.03	-0.12	-0.08	-0.07	0.18	0.03	0.10	-0.03	

CHAPTER V

DISCUSSION

A noteworthy advantage in using the Affinity 2.0 for this study was having direct access to our data. Having access to our sample's data was invaluable as it allowed us to calculate means, medians, *SD*, etc. which, consequently, allowed inter-individual comparisons to be made. Access to our data also provided the opportunity to conduct additional analyses, which allowed us to obtain valuable information regarding the response of our sample to the Affinity 2.0.

A Summary of Our Results

In short, the women in our sample demonstrated fairly inconsistent VT across categories. While the majority of the VT correlations were statistically significant, the practical significance of most of the PPMCC could be considered moderate at best, though the Spearman's Rho correlations began to approach practical significance. Thus, our sample's VTs for specific categories were less reliable than hoped. However, the correlations we discovered for our sample's raw ratings of categories were the highest correlations of all and possessed practical, in addition to statistical, significance.

Despite the lower VT correlations for specific categories, however, the existence of an overall discernable VT response pattern for normal heterosexual females was clear as our sample demonstrated identical categorical VT preferences at both test and retest (ADM, JUM, ADF, JUF, PJM, PJF, SCF, SCM). The observable preference for ADM and JUM images followed by ADF and JUF images with a clear drop in VTs for pre-juvenile and small child images was a clearly delineated pattern found throughout all parts of the data analysis (in means, medians, weighted mean ranks, and ratings). In addition, as assessed by a chi-square goodness of fit test, this VT pattern was found to be temporally stable from test to retest.

How Our Results Compare with Past VT Studies

Our findings regarding the VT pattern for normal heterosexual females was similar to the pattern reported by previous researchers (Quinsey et al., 1993; Wright & Adams, 1994; Quinsey et al., 1996). Like Wright and Adams (1994) and Quinsey et al. (1996), our sample of women rated members of the adult member of the opposite sex highest and looked longest at images of the preferred sexual stimulus. The VT pattern wherein they looked longest at images of the adult preferred sex stimuli with VT decreasing with the age of the model is also consistent with the findings of former studies (Quinsey et al., 1993; Wright & Adams, 1994; Quinsey et al., 1996). Hence, like past studies, our results confirm the existence of a distinct and discernable VT pattern for normal heterosexual women (Quinsey et al.,1993; Wright & Adams, 1994; Quinsey et al., 1996). However, unlike past studies that utilized only one trial, the present study was able to verify this existence of the VT pattern over two trials and to confirm the temporal stability of the VT pattern from test to retest.

Conversion of OTL to Ipsatized Weighted Ranks

The conversion of raw scores into weighted ranks was another advantage in utilizing the Affinity 2.0, as this conversion allowed for an assessment of temporal stability through a chi-square goodness of fit test. While the conversion from raw VT to weighted ranks was not exact, the transformation of our raw data into the Affinity 2.0 weighted ranks seemed to cause minimal distortion as the ipsative weighted rank response curve mirrored the OTL response curve and the integrity of curve was maintained as well as the orderings of preferred VT categories. Some loss of sensitivity was manifest as the mean weighted rank correlations yielded more conservative estimates than the raw score median VT correlations, with pre-juvenile and small child categories possessing the weakest correlations of all. In addition we also found that, on the whole, the women in our study were most consistent in their ranking of adult and juvenile categories and least consistent in their ranking images of pre-juvenile and small child categories.

While the conversion from raw VT to weighted ranks was not exact, the transformation of our raw data into the Affinity 2.0 weighted ranks seemed to cause minimal distortion as the ipsative weighted rank response curve mirrored the OTL response curve and the integrity of curve was maintained as well as the orderings of preferred VT categories. To help ascertain the degree and nature of the distortion that occurred in the transformation of this sample's OTL VTs to ipsatized weighted ranks, Brown (2005) conducted an in-depth investigation into various features (skew, standard deviation, variance, etc.) that contributed most to this distortion. He discovered that distortion was more likely to occur in cases where standard deviations in raw OTL VT for stimulus categories were larger, or when there was a greater variance in skew for raw VT. In other words, distortion is more likely to occur when greater variability existed in the time a person spent viewing each image in a category, or when a participant viewed one or two slides for a significantly longer time period than most of the other slides in the category (resulting in skew). For further explanation and discussion of the distortion that occurred in the transformation of this sample's raw OTL scores into ipsatized weighted ranks, refer to Brown (2005).

Chi-Square Goodness of Fit Test—Potential Future Uses

The chi-square goodness of fit test that was used in our study to ascertain the temporal stability of our sample's response from test to retest might also prove useful to researchers and clinicians in other capacities. Given that dissimulation is always a concern in assessment, the chi-square goodness of fit test could aid researchers and clinicians as a check for consistency in responding.

For example, in our study, we found both the VT and the rating curves to be temporally stable from test to retest. Given this fact, using a chi-square goodness of fit procedure to ascertain the temporal stability of self-report ratings in addition to a chi-square analysis of VT behavior could be an invaluable way to detect dissimulation in future administrations of the instrument as a discrepancy in the stability of either self-report ratings or VT could alert professionals to the possibility of a fraudulent or misleading response set. Given that Affinity 2.0 test data are readily and directly accessible to the user, these analyses would be both possible and easy to perform. In addition, the chi-square goodness of fit test also would allow professionals to make case-by-case comparisons. The following is an example to illustrate how a chi-square Goodness of Fit test might be used to determine if the results from a novel participant's to the Affinity 2.0 are significantly different than the response obtained in our study.

In this comparison, we can utilize the results from a participant who was not included in our study as she marked a category other than exclusively heterosexual on the Kinsey scale (Predominantly heterosexual with incidental homosexual). After obtaining her Affinity 2.0 Mean Rank scores, we can use the average response from our sample of 120 normal heterosexual women as the expected and participant 2008's data as the observed to determine if 2008's response is significantly different from that of our sample. The results of this analysis are shown in Table 16.

Table 16

Participant 2008 Comparisonc² Goodness of Fit Test

Category	ADM	JUM	PJM	SCM	ADF	JUF	PJF	SCF	TOTAL
Observed	27.57	23.29	29.43	25.86	24.07	36.79	24.00	37.00	228.01
Expected	43.61	36.55	25.95	20.00	29.99	27.79	21.73	22.29	227.91
O-E	-16.03	-13.21	3.43	5.76	-5.83	9.09	2.20	14.70	0.11
O-E SQ	256.96	174.50	11.76	33.18	33.99	82.63	4.84	216.09	
O-E SQ/E	5.89	4.78	0.45	1.65	1.14	2.98	0.22	9.69	26.81
Residual	-2.43	-2.19	0.67	1.28	-1.07	1.73	0.47	3.11	

A chi-square goodness of fit test was calculated comparing Participant 2008's response to the Affinity 2.0 to that of the average response of 120 normal exclusively heterosexual females. A significant result was obtained ($\chi^2(7) = 26.81$, $p < .001$). This suggests that participant 2008's response differs significantly from that found in our study. Residuals with an absolute value greater than 1.96 are considered to significantly contribute to our chi-square value. Upon examining Table 16, we find that participant 2008 has three values that exceed the 1.96 threshold—ADM (-2.43), JUM (-2.19), and SCF (3.11). Hence, we can conclude that participant 2008 looked at categories ADM and JUM for significantly shorter periods of time and looked at SCF for a significantly longer period of time than did our study's sample of 120 normal, exclusively heterosexual women.

As can be seen, in using a chi-square goodness of fit procedure, not only are we able to tell whether the novel case differed from our sample's overall response to the Affinity 2.0, we are also able to determine the categories in which we received a response that differed from what we would expect. This type of information could prove very helpful to researchers and clinicians and further research in the use of this procedure with Affinity 2.0 data seems warranted.

Elevations for ADF/JUF Images—The "Blip"

One of the more interesting findings in our VT pattern not discussed in previous studies was the presence of a "blip", or elevation, for the ADF and JUF categories. As this "blip" would not be anticipated with women who claim to be exclusively heterosexual it begs further investigation and explanation.

The "blip"—all about sexual interest? In a study recent by Israel and Strassberg (2006), VT was utilized as a measure of categorical (i.e., female vs. male) sexual interest. The participants were 51 self-identified heterosexual males and 55 self-identified heterosexual females who were undergraduate students (mean age 22, range 18-31, and mean age 21 years, range 18-40, respectively). Participants viewed 120 pictures on a computer presented in random order (via 2 trials of 60 pictures that included 25 pictures each of suggestive, clothed adult males and females and 10 neutral/landscape pictures). Viewing time was surreptitiously measured as participants rated the sexual attractiveness of each image on a 7-point Likert scale (where "1" is *Not At All Sexually Appealing* and "7" is *Extremely Sexually Appealing*).

Israel and Strassberg (2006) discovered that the VT patterns of men and women were sufficiently disparate to identify the participant's gender correctly over 90% of the time. A repeated measures ANOVA revealed a main effect for picture type, as well as an interaction between gender of participant and picture type. Specifically, it was reported that men looked at female pictures (*M* = 3.4 seconds) longer than neutral pictures (*M* = 2.1) and for twice as long as they viewed male pictures (*M* = 1.7). Women, on the other hand, viewed male pictures (*M* = 2.8) longer than either female (*M* = 2.6) or neutral pictures (*M* = 2.0).

When Sexual Appeal Ratings were analyzed using the same ANOVA design it was found that a main effect for picture type existed as well as an interaction effect between participant gender and picture type. Men rated pictures of females as significantly more sexually appealing (*M* = 5.4) then neutral or male pictures (*M* = 2.0, *M* = 1.4, respectively), while women rated pictures of males as more sexually appealing (*M* = 3.8), but not significantly more so than female (*M* = 3.1) or neutral (*M* = 2.0) pictures.

The researchers concluded that VT was a valid and reliable measure of categorical sexual interest as both women and men viewed opposite pictures for significantly longer than same sex pictures. In addition they mentioned, “of note, women’s between-category differences in viewing time were smaller than those of men, suggesting less category specificity for women” (Israel & Strassberg, 2006, October), meaning that “if you view gender as a category and sexual orientation as a preference for a gender category…we found that heterosexual…women…showed a moderate preference for males and also a moderate but lower preference to females” (E. Israel, personal communication, November 22, 2006). While it may be, as Israel and Strassberg (2006) maintain, that normal heterosexual women are also sexually attracted to other women, it seems that other potential explanations of this elevation in VT for female images might exist.

The ADF/JUF "blip"—another potential explanation. While the elevations in VT for the ADF and JUF categories in our study could potentially represent sexual interest, this would contradict the Kinsey scale self-report given by all of the participants. And, while it is conceivable that all of the participants in the study could have lied and failed to report sexual interest in other women, it is unlikely given that participants had no reason to dissimulate, results were kept confidential, and that numerous participants during the data collection process (not included in this study) did mark below the "Exclusively Heterosexual" option. Rather, it seems more feasible that VT may be a dual phenomena for women. In other words, the Affinity 2.0 may be measuring some other construct in addition to detecting sexual interest. One plausible alternative explanation for our sample's VT behavior with ADF and JUF slides may be that of social comparison.

Women's social comparison has been a phenomenon that has been observed and researched throughout various races—White, Black, Asian, Hispanic/Latina, etc. (Poran, 2002; Frisby, 2004; Evans, 2003; Alvarez, 2003). Many studies have highlighted the automatic, spontaneous, subconscious, and unintentional nature of these social comparisons and found that comparison is often the greatest when the image is similar to the individual (Frisby, 2004). Given that physical attractiveness (thin-ideal body type) is often associated with happiness and successful life outcome (Evans, 2003; Poran, 2002) and has been a means for women to gain power and privilege (Rudd & Lennon, 2000), the beauty standard holds more importance for women than men (Poran, 2002).

The importance/centrality of physical attractiveness for women was highlighted in a study conducted by Dijkstra and Buunk (2002). In this study, the 56 characteristics identified by Dijkstra and Buunk (2002) that were found to evoke jealousy in a potential rival for one's partner centered around five major factors: social dominance, physical attractiveness, physical dominance, social status, and seductive behavior. They discovered that, while men responded with jealousy to a rival's social status and physical dominance, women felt considerably more threatened and responded with significantly more jealousy than men when a rival was thought to be more physically attractive. More specifically, they found

> women feel more jealous than men when their rival has a more attractive face or body, has a better figure, has more beautiful legs, has a tighter waist, is better dressed, is more slender, is dressed more scantily, has more beautiful hips, [and] is built lighter. (p. 834)

It is interesting to note that none of these characteristics require interaction or any additional outside information for comparisons to be made as assessments of physical attractiveness are made visually.

Social comparison—a special case of social cognitions theory? Affinity 2.0 is based upon the idea that VT and sexual interest are a special case of Social Cognitions Theory. Like VT and sexual interest, social comparison might also be a special case of Social Cognitions Theory. With social comparison however, rather than determining whether a person/image is an exemplar of either "potential/desired sexual partner" or "not a potential/desired sexual partner," the distinction would be made as to whether the image/person is an exemplar of "potential rival" or "not a potential rival."

Consequently then, it is likely that a generalized prototype exists for those whom one would consider to be a potential rival or threat. An exemplar of the "potential rival" category would have to fit a generalized prototype of the appropriate sex and age. If the image/person fits this initial requirement, then the individual/viewer is likely to make either an upward or downward comparison based on more specific characteristics like body build, hair/eye color, race, etc. As suggested by the research, the more similar the person/image is to the viewer, the more likely a comparison will be made.

Determining how one compares to the model in question requires further cognitive processing which, in turn, results in more time being invested in the rating task. Given that the task in Affinity 2.0 is to provide a rating of the image solely based on physical characteristics that are assessed visually, social comparisons with similar others seem practically inevitable. Making this type of comparison would require additional time resulting in longer VTs for these images.

According to this hypothesis, we would expect that only ADF and JUF images would fit the generalized prototype as PJF and SCF images would meet the criteria for sex, but not for age (given the 18-56 year old age range of our sample). Because PJF and SCF images were not similar enough in age to our sample to qualify for a "potential rival," no upward or downward comparison would be made, and, thus, less VT would elapse.

Conversely, ADF and JUF images would fill the generalized prototype requirements for a "potential rival." Thus, we would expect the participant to make an upward or downward comparison with the image depending on factors such as body build, hair/eye color, race, etc. The more similar the image is to the participant the more we might expect the participant to compare. Given that models of numerous ethnicities and races are portrayed in the images shown in Affinity 2.0, it is plausible that every participant could find at least one image with which she was similar enough to compare herself to. In addition, considering that the images in Affinity 2.0 are of models who are attractive and typically represent the thin-body ideal, the possibility for these images to qualify as targets for social comparison is high.

Hence, overall, we would expect lower VTs for PJF and SCF images and higher VTs for ADF and JUF images. Indeed, this is what was observed in our sample. Given our results, social comparison seems to be a good candidate for explaining the elevated VTs for ADF and JUF images in a self-reported exclusively heterosexual female sample.

Sexual interest, social comparison, or other?—The necessity of further VT studies. While we can merely hypothesize about what the ADF/JUF blip is about, conducting a similar study with normal heterosexual females that incorporates a qualitative component would be more considerably more helpful in resolving this unknown. After viewing and rating the images, participants would then be interviewed about the images that they lingered on for the longest amount of time. While this change in procedure would reveal to participants the actual, previously surreptitious, focus on VT behavior, it seems tolerable as the information provided in these discussions is perhaps the only way for us to discern with any certainty the meaning of this elevation in VT for ADF and JUF slides. This information would be invaluable in helping us to ascertain whether this blip in the VT pattern is indeed due to sexual interest, to social comparison, or to another phenomenon (e.g., an aesthetics/beauty effect).

If indeed we discover that this “blip” is due to social comparison, or other phenomena, it will then have to be determined to what degree this competing construct is a confounding factor in the use of the Affinity 2.0 as a tool for measuring sexual interest. In other words, if the VT patterns observed with women are due to a number of phenomena rather than primarily sexual interest (as has been hypothesized), then the ability of researchers to utilize VT, and consequently, instruments like the Affinity 2.0, to provide an accurate measure and classification of sexual interest and preference for women would be compromised at the least and erroneous and invalid at worst.

Potential Future Studies with the Affinity

Previous studies (Quinsey et al.,1993; Wright & Adams, 1994; Quinsey et al., 1996) have used various comparative populations—heterosexual males/females and homosexual males/females—to determine how unique and distinguishable VT patterns were among the different groups. Future studies with the Affinity 2.0 that included normal heterosexual males, normal homosexual males/females, normal adolescent males/females, etc., as comparative populations would help researcher to further discern how distinct the VT pattern for normal heterosexual women and other groups are (e.g., are VT patterns distinct enough to discriminate between members of various groups), would allow for comparison, and would provide VT patterns which researchers could use to compare novel cases against (i.e., to identify abnormal VT responses).

Limitations and Suggestions

In retrospect, and after an examination our data, it is plausible that the period of two weeks between test and retest may not have been sufficiently long enough given the fairly consistent drop in VT of around one second from time one to time two. This consistent drop in VT may make responses seem more unstable than they really are. It is supposed that this phenomena (i.e., the one second decrease in VT) may be due to a sort of recency, familiarity, recognition, or recall effect, and thus could be corrected in future studies through utilizing a longer delay between test and retest. In our study we also had a problem with misratings. Providing clarification about the rating system (i.e., reminding participants to rate how sexually attractive the image is to them *personally*) at both test and retest and/or adding a component to the standardized instructions wherein participants are asked to repeat the basis of rating images back to the assessment administrator could help prevent this problem in future studies.

The fact that this study was conducted on a religious, sexually conservative campus might also be a potential limitation as some participants may have been uncomfortable with the task, are not as likely to have experimented with sexuality (given that sexual activity outside of marriage is prohibited as a part of the Honor Code and the majority of our sample were single), or may have felt an expectation to identify as exclusively heterosexual. It is impossible to know if how much of an impact this may have had. It also seems as if this factor would be difficult, if not impossible, to control for. Conducting a similar study elsewhere and comparing the results would be interesting and might help to determine how our results differed, if there was a significant difference. In addition, performing a study within the same BYU population wherein only married (i.e., sexually experienced persons) were included might also be an interesting study to which we could compare results.

Another limitation related to our sample is that of limited diversity. In addition to the participants' self-reports of being exclusively heterosexual, our sample consisted primarily of Caucasian, highly educated, young, religious women. Thus, generalizations of these results to any person or group that differs significantly from this population should be done with great caution, if at all. Future studies with differing multicultural populations are also recommended.

A suggestion for the Affinity 2.0 itself would be to add more images to the assessment. Given that the current version of Affinity contains only seven slides for each category, an increase in the number of images in each category would likely help to increase the reliability of the instrument overall. Evidently, for this very reason, the new version of the Affinity—the Affinity 2.5—now has ten slides within each stimulus category (personal communication, Lane Fischer, November 30, 2006).

Conclusion

In conclusion, our study affirmed the existence of a distinct and distinguishable VT pattern for normal heterosexual females at both test and retest that is consistent with findings from previous researchers. In addition, due to the ability of the Affinity 2.0 to convert raw OTL VT scores into ipsatized weighted ranks with minimal distortion, a chi-square goodness of fit test was able to be performed that affirmed the temporal stability of this aforementioned VT pattern. The use of the chi-square goodness of fit procedure for purposes including detection of dissimulation or case-by-case comparisons seems promising.

One particular finding of this study that begs further exploration and explanation is the presence of an ADF/JUF blip in the VT pattern for normal heterosexual women. While other researchers insist this elevation is simply an indication of a lower but moderate sexual preference for women, it is unclear whether this is the case or not as other explanations (e.g., social comparison) seem feasible. A future study that includes a qualitative component would help provide this needed clarification.

Thus, while many aspects of the Affinity 2.0 seem promising, further studies are necessary to determine whether the Affinity 2.0 can provide the “accurate measure and classification of sexual arousal and preference” researchers and clinicians are searching for.

REFERENCES

Abel, G. G., Huffman, J., Warberg, B., & Holland, C. L. (April, 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.

Alvarez, R. (2003). The effects of acculturation and internalization of the White American beauty ideal on social comparisons, self-esteem, and body satisfaction among adult Mexican American women [Abstract]. *Dissertation Abstracts International: Section B: The Sciences and Engineering, Vol 63 (9-B),* 2003, 4359.

Bailey, J. M., Gaulin, S., Agyei, Y., & Gladue, B. A. (1994). Effects of gender and sexual orientation on evolutionary relevant aspects on human mating psychology. *Journal of Personality and Social Psychology, 66* (6), 1081-1093.

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

Benson, E. (2003). The science of sexual arousal. *Monitor on Psychology, 34* (4), 50-52.

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

Brown, A. (2005). *Ipsative score distortion on Affinity 2.0.* Unpublished master's thesis, Brigham Young University, Provo, Utah.

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.

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

Dijkstra, P. & Buunk, B. P. (2002). Sex differences in the jealousy-evoking effect of rival characteristics. *European Journal of Social Psychology, 32*, 829-852.

Evans, P. C. (2003). "If only I were thin like her, maybe I could be happy like her:" The self-implications of associating a thin female ideal with life success. *Psychology of Women Quarterly, 27,* 209-214.

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., & Morgan, D. (2006, September). *Norm-referenced clinical decision-making with Affinity viewing-time results*. In L. Fischer (Chair), *Current Research and Clinical Application of Affinity Viewing-Time Scores.* Symposium conducted at the 25th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL.

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.

Frisby, C. M. (2004). Does race matter? Effects of idealized images on African American women's perceptions of body esteem. *Journal of Black Studies, 34 (3)*, 323-347.

Glasgow, D. (2003). Affinity 2.0 [Software and Instruction], David Glasglow: 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.* Symposium conducted at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, FL.

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.

Heiman, J. (1977). A psychophysical exploration of sexual arousal patterns in females and males. *Psychophysiology, 14*, 266-274.

Heiman, J. R. (1980). Female sexual response patterns: Interactions of physiological, affective, and contextual cues. *Archives of General Psychiatry, 37*, 1311-1316.

Hoon, P. W. (1984). Physiologic assessment of sexual response in women: The unfulfilled promise. *Clinical Obstetrics and Gynecology, 27* (3), 767-780.

Howell, D. C. (2002). *Statistical methods for psychology 5th Ed.* Pacific Grove, CA: Duxbury.

Israel, E. & Strassberg, D. S. (2006, October). *Viewing time as an objective measure of sexual interest in heterosexual men and women.* Poster session presented at the annual meeting of the Association for the Treatment of Sexual Abusers, Chicago, IL.

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.

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* (4), 624-629.

Macrae, N. & Bodenhausen, G. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology, 51*, 93-120.

Mandell, R. (N.D.). Evaluating shorter versions of the Marlowe-Crown Social Desireability Scales. National Institutes of Health. Retrieved on 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* (1), 162-199.

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

Poran, M. A. (2002). Denying diversity: Perceptions of beauty and social comparison processes among Latina, Black, and White women. *Sex Roles, 47 (1/2)*, 65-81.

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.

Rudd, N. A. & Lennon, S. J. (2000). Body image and appearance-management behaviors in college women. *Clothing and Textiles Research Journal, 18 (3),* 152-162.

Singer, B. (1984). Conceptualizing sexual arousal and attraction. *The Journal of Sex Research, 20* (3), 230-240.

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 visual 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, homogenous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191-193.

Symons, D. (1979). *The Evolution of Human Sexuality*. New York: Oxford University Press.

Wincze, J. P., Hoon, E. F., & Hoon, P. W. (1978). Multiple measure analysis of women experiencing low sexual arousal. *Behavioral Research and Therapy, 16*, 43-49.

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

Informed Consent

Consent to be a Research Subject Participant # 2XXX

We would like to request your participation in a study designed to investigate a newly developed instrument that purports to measure sexual interest. 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. <u>No pornographic images are used in this study.</u>

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.

Kara Harmon
(801) 687-5272
kara_harmon@byu.edu

Lane Fischer, Ph.D.
(801) 422-3857
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 B

Affinity 2.0 Prototype Line Drawings and Sample Images

Image 1

Affinity 2.0 Prototype Line Drawings

(Glasgow, 2003).

Image 2

Affinity 2.0 Sample Images

(Glasgow, 2003)

APPENDIX C

Demographics, Social Desirability and Sexual Interest Questionnaire (DDSQ)

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