2011-07-13

Reliability and a Measure of Sexual Interest: Examining the Temporal Stability of Scores on Affinity 2.5

Kristina S. Withers Hansen
Brigham Young University - Provo

Reliability and a Measure of Sexual Interest: Examining the Temporal Stability of Scores on Affinity 2.5

Kristina S. Withers Hansen

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Lane Fischer, Chair
Rachel Crook Lyon
Aaron P. Jackson
Steven A. Smith
Ellie L. Young

Department of Counseling Psychology and Special Education

Brigham Young University

August 2012

ABSTRACT

Reliability and a Measure of Sexual Interest: Examining the
Temporal Stability of Scores on Affinity 2.5

Kristina S. Withers Hansen
Department of Counseling Psychology and Special Education, BYU
Doctor of Philosophy

Affinity 2.5 is a computer-based instrument designed to measure sexual interest using viewing time of images depicting fully-clothed males and females of different ages. Participants are asked to rate the sexual attractiveness of the person in the image according to a 15-point scale while their viewing time of each image is surreptitiously monitored. The validity of viewing time as a measure of sexual interest is based on social cognition theory and is established in the review of literature.

The number of images comprising Affinity 2.5 represents a 42.9% increase from the previous version of the assessment, Affinity 2.0. The purpose of this study was to examine the temporal stability of scores on Affinity 2.5 for a sample of exclusively heterosexual, nonpedophilic males and females. Viewing time data from 63 males and 84 females were analyzed using a chi-square procedure. Results of this analysis indicate that 86% of responses from the male participants and 88% of responses from the female participants were consistent from time one to time two. As suspected, these percentages represent an increase in reliability over the temporal stability of the shorter Affinity 2.0.

Keywords: Affinity 2.5, sexual interest, viewing time, reliability, temporal stability, social cognition theory, chi-square, David Glasgow, Lane Fischer.

ACKNOWLEDGMENTS

Sincere appreciation to my graduate committee for their patience and support through this lengthy process, with special recognition and deepest gratitude to my graduate advisor and dissertation chair, Dr. Lane Fischer. Lane's direction, encouragement, and understanding of both me and my writing process have changed for the better my beliefs about myself and the world around me. His impact has been immeasurable and his company and friendship ever enjoyable.

Thanks to the members of the Affinity Team for their hours of data collection and idea generation, and for their motivation for progress on this project.

Special thanks to my dear cohort friends for their invaluable support as we worked together toward completion and defense of our dissertations. Their joining me in laughter and tears through our precious years together brought richness and vitality to relationships more meaningful than I could have imagined.

Most importantly, I am thankful for my husband and best friend, Justin, who has been with me through it all. He is my favorite part of life, and for the person he is I am truly grateful.

TABLE OF CONTENTS

APPENDICES

TABLES

FIGURE

Introduction

The prevalence of criminal adult and juvenile sexual offending and victimization demands improved methods of assessment of sexual interest and studies of sexual deviance. Data reflecting the number of sexual crimes against adults and children are sobering. For example, Finkelhor, Ormrod, Turner, and Hamby (2005) report that by age 17, one in twelve youth have been sexually victimized. Among juvenile victims of violent crimes, three in four females and one in four males were victims of sexual assault (USDOJ, retrieved 1/18/2010). According to data collected in the United States for the year 2007 by the Federal Bureau of Investigation (the most current data), there were 90,427 forcible rape offenses reported (U.S. Census Bureau, 2011). The United States Department of Justice *Juvenile Offenders and Victims: 2006 National Report* indicates that juveniles are the victims in 70% of sexual assaults.

The majority of sexual offenses are perpetrated by males (Snyder, 2002), but females also commit sexual crimes. Snyder and Sickmund (2006) report significant increases in female-perpetrated sex offenses from 1997 to 2002. Meta-analyses of sexual recidivism data indicate that one of the strongest predictors of repeat offenses is deviant sexual interests (Hanson & Bussière, 1998; Whitaker et al., 2008), or "enduring attractions to sexual acts that are illegal (e.g., sex with children, rape) or highly unusual (e.g., fetishism, autoerotic asphysia)" (Hanson & Bussière, 1998, p. 2). According to these authors, sexual deviance can be conceptualized both as sexual behavior that is illegal and sexual interests that deviate from socially and legally acceptable norms.

Intervention and treatment programs for sex offenders are available, but accurate and reliable assessment is a crucial precursor to clinical intervention (Wright & Adams, 1994). Current sexual interest assessment instruments include analysis of records, clinical

interviews/self-report, polygraph testing, penile and vaginal plethysmography, and viewing time measures.

Clinical interviews and other self-report measures are subjective tools useful for gathering information about a person's sexual interests. These measures pose various problems, however, due to the sensitive nature of discussing sexuality, the possibility for dishonesty or dissimulation, the difficulty of identifying the correct and most useful questions to ask, and the desire of sexual offenders and participants in sexuality studies to appear normal (Gress, 2005; Meston, Heiman, Trapnell, & Paulhus, 1998; Wincze, Hoon, & Hoon, 1978; Wright & Adams, 1994). Plethysmography grew out of a need for an alternative to self-report measures and has been widely used to assess sexual arousal in both males and females. Admittedly, penile and vaginal plethysmography has its limitations, but several findings suggest that genital response is a valid measure of sexual arousal (Howes, 2003; Letourneau, 2002). Unfortunately, as strictly a measure of sexual arousal, plethysmography is not necessarily a measure of sexual preference or of risk of offending. Research suggests that concerns about the inadequacy of assessment of genital response demand another, more adequate measure of sexual interest (Gress, 2005; Fischer, 2000; Marshall, 1996).

Instruments using viewing time to measure sexual interest are becoming increasingly accepted and empirically validated. Viewing time instruments are based on the theory that individuals will look longer at something to which they are attracted. Social cognition theory of sexual interest suggests that viewing time increases when the stimulus image closely fits the individual's ideal of attractiveness because more cognitive processing is necessary to categorically define the object (Macrae & Bodenhausen, 2000; Macrae & Bodenhausen, 2001).

Currently there are two viewing time instruments used to assess sexual interest. One is the Abel Assessment for Sexual Interest (AASI), and the other is the Affinity. This study uses

Affinity 2.5, the most recent viewing time assessment of sexual interest. Wright and Adams (1994) explain that 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" (p. 221). If assessment tools are being used to inform clinical and legal decisions that affect a person's life, it is important that the information provided by the instrument be reliable, have evidence of validity, and be interpreted correctly.

Statement of Problem

As Fischer (2000) explains, exploring an instrument's test-retest reliability is crucial to the development and validation of any assessment tool that purports to measure a construct considered to be relatively constant across time. Sexual interest is one such stable construct. Reliability is defined as "consistency of measurement results" (Warner, 2008, p. 830) and sets the limits on validity. A reliable instrument measuring a stable construct would be expected to produce highly similar results under different administrations of the same test to the same person. It is important to study and illuminate the psychometric properties of an instrument before its data is used to make decisions that affect people.

Affinity 2.5 is a revised and expanded version of a computer-based assessment tool used to measure sexual interest. Affinity 2.5 uses 42.9% more slides than the previous version of the Affinity. These additional stimulus images seem to represent a significant change in the instrument. The temporal stability of Affinity 2.5 responses from nonpedophilic, exclusively heterosexual males and females is unknown.

Statement of Purpose

The purpose of this study is to examine the temporal stability of scores on Affinity 2.5 for a sample of nonpedophilic, exclusively heterosexual males and females.

Review of Literature

Current efforts to assess sexual interest and sexual deviance include analysis of records, clinical interviews, self-report, penile and vaginal plethysmography, and viewing time measures (Cloyd, 2006; Crosby, 2007; Fischer, 2000; Harmon, 2006; Laws, 1989; Laws & Gress, 2004; Marshall, 1996). While each of these various methods has its merits and flaws, the method with which this study is concerned uses visual attention as a measure of sexual interest. This review of relevant literature begins with evidence of child and adult ability to perceive physical attractiveness in others, an examination of this perception's relationship to social cognition theory, and the theory's connection to viewing time. A historical review traces across time the progression of research using different viewing time measures of sexual interest. The review of literature includes discussion of external influences on viewing time and a description of current viewing time measures of sexual interest and their component strengths and weaknesses. Temporal stability is defined with application to the current study.

Sexuality research regarding measures of sexual interest, assessment tool reliability and validity, developmental issues, ethical concerns, definitions of terms, normative samples, and standardization is ongoing and expanding. Incidentally, it is somewhat surprising that for a construct so innately human as sexuality, the research into the assessment of sexual interest is not more voluminous. Nevertheless, it is important to understand the rationale and research behind previous empirical efforts to develop and utilize measures of sexual interest. Such efforts and their results are salient for studying and establishing sustained visual attention as a measure of sexual interest

Visual Perception of Physical Attractiveness

Visual perception of physical attractiveness is an important aspect of humanity related to sexual interest and attraction. In a field of research known as "esthetics," studies during the

1950s and 1960s were conducted to examine cultural and personality variables in esthetic and color preferences (Child, 1962; Child & Cooperman, 1968; Child, Hansen, & Hornbeck, 1968; Child & Kasti, 1968). Other attempts have been made to relate theoretical esthetics of art and beauty to psychological issues (Valentine, 1962; Arnheim, 1954; Arnheim, 1966). In their seminal (and perhaps pessimistic) review of physical attractiveness effects, Berscheid and Walster (1974) concluded that there was no answer to the question of what constitutes beauty. In a manner typical of challenge to broad, general-sounding statements such as Berscheid and Walster's, efforts to expand and define physical attractiveness of the human form grew out of this early research.

Children as judges of physical features. Studies involving children as judges of physical attractiveness have suggested general consensus among participants (as demonstrated by high interrater reliability) regarding criteria of physical attractiveness.

Body attractiveness. Even young children have shown evidence of preferences for particular body characteristics when presented with visual stimuli. Specifically, Staffieri (1967) found that a preference for the male mesomorphic body type (muscular or sturdy body build) became evident as early as age six, but did not become significant until about age eight. Similarly, in a 1973 study by Cavior and Lombardi, these researchers used a reliability study to conclude that the "cultural criteria used by older persons begin to be acquired at age 6 [*sic*]" (p. 69). In particular, the interrater reliability of physical attractiveness rankings for images of 11- and 17-year-olds from male and female judges ranging in age from five to eight ($N = 62$) reached statistical significance among the six-year olds, and increased among the seven- and eight-year-olds. Interestingly, the eight-year-old judges ranked the attractiveness of the adolescent persons pictured in the images very similarly to the way older comparison groups (ages 10 to 17) ranked the images of their peers. In addition, among judges aged seven and eight years, no differences

were found between the male and female judges' rankings of the physical attractiveness of the 11- and 17-year-olds. The results from both of these studies suggest that children are remarkably adept at discerning physical attractiveness when presented with photographic stimuli, that such discernment is fairly similar for children of the same age, and that visual perception of physical attractiveness develops quite early in life.

Facial attractiveness. A few years later, Dion (1973) used a sample of 61 boys and girls ranging in age from three years to six years to "assess whether young children exhibit stereotyping based on facial attractiveness" (p. 183). The preschoolers were shown facial photographs of peers who, by adult standards, would be considered attractive or unattractive. Using chi-square analyses, Dion found that subjects reliably discriminated differences in facial attractiveness, and that the preschoolers' judgments were in the same direction as adults' judgments. There were no effects for participant gender. While the results of these several studies differ regarding the age at which judgment of physical attractiveness becomes reliable, it is safe to conclude that even very young children develop opinions of physical attractiveness, and that these opinions tend to be similar to adult opinions.

Several studies during the 1980s concluded that even infants (ages three- to six-months) appear to have opinions about attractive and unattractive faces. For example, infants who were shown pictures of adult-judged attractive and unattractive faces appeared to prefer attractive ones (Langlois, et al., 1987; Langlois, Roggman, & Rieser-Danner, 1990; Samuels & Ewy, 1985; Shapiro, Eppler, Haith, & Reis, 1987). Langlois and Roggman (1990) suggest that "even before any substantial exposure to cultural standards of beauty, young infants display behaviors that seem to be rudimentary versions of the judgments and preferences for attractive faces so prevalent in older children and adults" (p. 115). According to these researchers, it seems possible, if not probable, that there exist among infants, older children, and adults certain

predominant stimulus dimensions of faces that are considered attractive; and that the ability to detect these stimulus dimensions may be innate or acquired earlier than previously believed.

Adults as judges of physical features. Studies of adults' perception of physical attractiveness indicate that the appearance of certain physical characteristics may be more important than others.

Body attractiveness. In a 1981 study by Horvath of physical attractiveness ratings of front-view line drawings of male and female physiques, the researcher found a high degree of consensus for what constitutes attractiveness among the study participants. Specifically, Horvath found that both males and females rated female physiques with greater curvature as less attractive. Male subjects' ratings were unaffected by breast size, while female subjects showed slight negative evaluation of large breasts. Both males and females rated broad shoulders, greater chest muscularity, and a slim waist as attractive in male physiques. It is interesting to note that even in a study using only line drawings, as opposed to actual photographs, both males and females tend to agree on which physical features contribute to increased attractiveness.

Hassebrauk (1997) conducted an interesting study of college-age male and female participants' ($N = 70$) viewing sequence, viewing time, and amount of information used with respect to various face and body parts in judgments of physical attractiveness. Participants were given the express task of using the least possible visual information to judge the attractiveness of target persons as depicted in images of 10 males and 10 females wearing swimming attire. Participants were not informed that their viewing time was being monitored and recorded. According to Hassebrauk, the results indicate that features associated with the mate value of the individual were looked at sooner and more often by both sexes. Specifically, for images of females, features that indicated youthfulness, sexual maturity, and fertility were viewed longer and earlier in the sequence of visual attention; and for images of males, features that suggested

status and dominance were looked at sooner and more often. This researcher (1997) also concluded that while the physical attractiveness of females appears to be more complex than that of males, his study clearly shows that facial attractiveness plays a much larger role than the attractiveness of the body in judgments of overall physical attractiveness.

Facial attractiveness. Another study of facial attractiveness by Cross and Cross (1971) used male and female participants of both African American and European American ethnicities ranging in age from seven years to adult ($N = 300$) as raters according to a seven-point scale of the perceived beauty of faces as depicted in photographs. Results of this study showed that females of both races down-rated adult male faces. Overall, female faces and adolescent faces received higher ratings than other sex and age groups, and African Americans gave higher ratings of beauty than did European Americans. The authors of this study concluded that male and female persons from two different ethnicities demonstrated an ability to detect beauty in faces of persons differing markedly from themselves in age, sex, and race.

Mueser, Grau, Sussman, and Rosen (1984) conducted two studies of physical attractiveness using college age males as participants. Photographs of 15 female target persons posing happy, neutral, and sad facial expressions were rated for facial attractiveness using paired comparisons and Likert scales. Two hundred males participated in these ratings. Results showed that target persons were rated as less attractive when posing sad expressions than when posing neutral or happy expressions. The ratings of the neutral or happy expressions did not differ significantly from each other. In the second study, 21 male college students rated on a 10-point Likert scale photos of a whole person, just the face (neutral expression), or just the body (from the neck down) of the 15 target persons used in the first study. Multiple regression analyses of the ratings revealed that still photos of the face contributed slightly more to overall attractiveness than did photos of the body (F-test for face $F(1, 13) = 7.16$, $p < .03$; F-test for body $F(1, 13) =$

5.19, $p < .05$). Facial attractiveness accounted for 27.56% of the total variance in overall attractiveness, and bodily attractiveness accounted for 19.97% of the total variance.

Other studies have focused on specific features of the face and their relative correlations to ratings of physical attractiveness and specific personality variables. For example, Cunningham (1986) used a sample of college-age males ($N = 75$) and 50 black and white photographs of female faces to study which particular facial configurations might be judged as intrinsically attractive. Cunningham found a positive correlation between attractiveness ratings and neonate features of large eyes, small nose, and small chin; maturity features of prominent cheekbones and narrow cheeks; and expressive features of high eyebrows, large pupils, and large smile. Extensions of the ratings by these participants also predicted personality attributions (of the people pictured), altruistic inclinations for the people pictured (by the participants), and participant reproductive interest (in the people pictured). According to the results of this part of the study, persons with more attractive features were more likely to be chosen for self-sacrificial and physically risky actions (study participants indicated that they would be willing to do something risky for the person pictured), for a job, dating, sexual preferences, and childrearing, although not for monetary investments. Similarly, Berry and McArthur (1985) used adult male faces as stimulus images for 40 male and 40 female participants. The study participants rated specific facial features and personality dimensions of physical attractiveness, age, and babyfacedness (large, round eyes, high eyebrows, small chin) for each image. Berry and McArthur found that adult males who were perceived as babyfaced were also perceived as more honest, naïve, kind, and warm than their more mature-faced peers. These studies suggest that not only are people capable of visually discerning physical attractiveness, but that they hold certain perceptions about what they see.

Physical Attractiveness, Social Cognition Theory, and Viewing Time

The previously discussed literature sought to explore the ability of humans to perceive physical attractiveness in others. As this line of research expanded, studies began to include questions of meaning regarding perception of physical attractiveness, and how such perception is used to make sense of the world. In 1977, Adams conducted a review of physical attractiveness research relating outer appearance and inner psychological characteristics. While the bulk of Adams's review is beyond the scope of the current review of literature, his conclusions are related to the idea that perception of physical attractiveness is a skill that is innately human. For example, Adams states that there is evidence to suggest "individuals hold a common standard of attractiveness which they use to evaluate the physical appearance of others" (p. 231). Social cognition theory is one way of understanding this innately human way that people attempt to make sense of the people around them. Social cognition researchers describe how perceptual cues, particularly distinctive visual cues (Zebrowitz, 1996), cause individuals to categorize other people and to imbue them with the properties that are described by the person's schema of that group (Fiske & Taylor, 1991; Hamilton & Sherman, 1994).

Macrae and Bodenhausen (2000) maintain that "in attempting to make sense of other people, perceivers regularly construct and use categorical representations to simplify and streamline the person perception process" (p. 93). These categorical representations may be conceptualized as "schemas," (Bartlett, 1932), or everyday theories that shape how people view and use information (Howard & Renfrow, 2003). Schemas are a type of system of categorization on which humans rely in order to process efficiently incoming information (Zerubavel, 1996). Zerubavel describes this categorization as a process of "lumping" and "splitting," where similar information is lumped together into mental clusters, and dissimilar information is split into distinct categories. People cognitively represent physical or social categories as schemas that

describe the attributes of the category and the relationships among those attributes. Category schemas vary from concrete exemplars of the category to abstract, fuzzy sets of loosely related attributes, or prototypes (Rosch, 1978).

Person schemas, for example, organize knowledge about particular types of people, generally emphasizing traits, such as physical or sexual attractiveness, or personality categories (Howard & Renfrow, 2003). Some theorists posit that the entire process is automatic, and others propose that portions of social cognition are deliberate. Allport (1954) explained that "the human mind must think with the aid of categories . . . We cannot possibly avoid this process. Orderly living depends upon it . . . Every event has certain marks that serve as a cue to bring the category of prejudgment into action" (p. 21). As an unequivocally automatic mental event, Allport construed as unavoidable such categories of prejudgment.

In contrast to Allport's claims of categorization automaticity, Fiske and Neuberg (1990) argue that sometimes people use handy cognitive shortcuts, specifically schemas, in judging others, but at other times, they pay close attention to the data and process them carefully. Macrae and Bodenhausen (2001) argue that such categorical thinking is individual—as opposed to universally automatic—and is affected by meaningful variation in the frequency and consistency of people's exposure to stereotypes and evaluations associated with their categories. One effect of social categorization is that it causes the perceiver to accentuate similarities among members of the same category and differences between categories. This appears to be a general consequence of categorization (Tajfel, 1969, as sited in Howard & Renfrow, 2003), but one that is asymmetrical because people tend to see out-groups as more homogeneous than in-groups. A popular explanation for this asymmetry is that we are more familiar with the in-group and therefore have more individuating information about in-group members than we do for out-group members (Linville et al., 1989). It seems logical, then, that activation of additional cognitive

structures regarding specific characteristics of in-group persons would be required in order to make judgments about individuals in that group.

Consider, for example, an explanatory scenario wherein an individual is presented with an image of a person and asked to decide if that person is sexually attractive to him or her. Under Macrae and Bodenhausen's (2001) social cognition logic, the person will activate schemas in order to determine the person's in-group or out-group status, and then employ his or her own categorical thinking based on meaningful factors associated with the social meaning of the stimuli—in this case, the pictured person's sexual attractiveness. Worsham (2009) explains:

> If the person in the image does not fit the individual's category of preferred sexual partners, the decision is made and viewing stops. However, if the image does fit this individual's conception of sexual attractiveness, new categories are activated based on specific traits that make the person in the image a category exemplar or a good fit in that schema. This process of category activation requires more cognitive work and thus takes more time. (p. 11)

According to this theory, once a stimulus is perceived as physically attractive, additional categorical cognitions might be employed to determine if the stimulus is sexually attractive, a meaningful social category. If the stimulus fits the individual's category of sexual attractiveness, further cognitions, stereotypes, and meanings may be considered in order to determine the stimulus image's degree of sexual attractiveness. Such considerations take time and would likely increase an individual's viewing time of the stimulus image while category determinations are being made. Under this logic, sexual interest might be considered a mediating precursor to decisions about sexual attractiveness since an individual must be sexually interested enough in a stimulus image to determine the degree to which it is sexually attractive.

Sustained Visual Attention as a Measure of Sexual Interest

Social cognition theory allows for timed sustained visual attention to stimuli to be used as a measure of sexual interest. In 2003, Glasgow and Croxen explored cognitive processing and category activation in relation to viewing time of stimulus images. These researchers (and others) have improved unobtrusive viewing time measures based on the idea that evaluation of sexual attractiveness requires more category activation and cognitive processing, which lead to longer viewing times. Similar efforts to employ viewing time as a measure of sexual interest and sexual attraction began in the mid-twentieth century. Based on the idea that viewing time offers a direct quantitative assessment of interest in sexual stimuli that does not rely upon subjective reports or genital response (Rupp & Wallen, 2009), continued developments in this field explore different aspects of the sexual interest construct.

Research beginnings. Rosenzweig (1942) developed an instrument called a "photoscope" intended to measure sexual interest in response to hormone therapy. Participants in his study included 20 inpatient males who had been diagnosed with schizophrenia. The participants were divided into two groups according to high and low interest in sexual behavior as rated by hospital personnel regarding the presence or absence of the patient demonstrating masturbation, heterosexual behaviors, and/or homosexual behaviors. The photoscope used two types of visual stimuli: 10 sexual and 10 nonsexual photographs mounted on cards set into a Rolodex-type apparatus so that they could be displayed one at a time in a box. Each patient controlled the length of time he viewed each photograph while an experimenter covertly timed how long each photograph was displayed. Patients participated in three sessions of the same procedure. The scoring of the photoscope test results was based on the principle that "the extent of sexual interest would be reflected quantitatively in the comparative difference in time spent on sexual and nonsexual pictures" (p. 152).

Viewing time results from each session of Rosenzweig's study differentiated between the high sexual group and the low sexual group by the greater length of time that the high group viewed the sexual over the nonsexual pictures. For example, during the first viewing session the high group spent a mean viewing time of 19 seconds on the nonsexual pictures, and a mean viewing time of 40 seconds on the sexual pictures. The low group viewed the nonsexual and sexual pictures for mean times of 19 and 13 seconds, respectively.

Conclusions regarding interpretations of this data are tenuous. Rosenzweig noted that it is possible that the low group's shorter viewing times of sexual stimuli may not actually reflect an uninterest in sex, but rather evidence sexual inhibition; in which case the shorter viewing times of sexual stimuli might represent the patients' efforts to protect themselves from demonstrating sexual interest for any number of reasons. Even so, Rosenzweig concluded that the results of this study appear to distinguish the two groups of men according to their viewing time of sexual stimuli, and establish sufficient reliability and validity of the photoscope for use as one of several measures of sexual interest. Rosenzweig conceded, however, that without a baseline, or normative sample with which to compare his results, it might be inappropriate to use his study of viewing time as a measurement of absolute sexual interest.

Research expansion. Several years later, Zamansky (1956) judged the results of his study to support the assumption that "object choice will manifest itself in the pattern of an individual's visual fixations, if these fixations can be measured without his awareness" (p. 445-446). Apparently this author believed that sustained visual attention reflected sexual interest, too. Zamansky used a sample of 20 homosexual males and 20 heterosexual males to covertly compare viewing times of paired photographs of male, female, or neutral content (landscapes). Participants were asked to determine which of the paired item photographs was bigger in size while their eye movements and viewing latency were tracked. The researcher expected that

homosexual men would spend relatively more time than heterosexual men observing photographs of males over females, males over neutrals, and neutrals over females. The results of this study verified each of these expectations and significantly discriminated the homosexual group from the heterosexual group.

Bullock (1959) obtained results similar to Zamansky's with a study demonstrating that the sexual content of a stimulus differentially affects the duration of attention in subjects of different genders. Specifically, when male and female participants viewed pictures of nude females and landscapes, females were less attentive to the nude females than to the landscapes, and males were more attentive to the nude females than to the landscapes.

In 1966 Leckart, Keeling, and Bakan conducted a viewing time study involving 27 male and 25 female college students that yielded results contrary to expectations. Like their sexual interest and viewing time research predecessors, these authors hypothesized that heterosexual participants "would spend relatively more time attending to pictures of members of the opposite sex than to pictures of their own sex" (p. 374). The research stimuli were 40 black and white photographs of adult males or adult females. Study participants were instructed to look at each photograph, one at a time, for as long as he or she liked. Participants were also told that they would not be asked to remember any of the pictures or anything about them. The time spent viewing each photograph was recorded. Individual comparisons revealed that females looked significantly longer at female pictures than at male pictures ($t = 2.91$, $df = 24$, $p < .01$), but that males did not differentially attend to the male and female pictures. A couple of extraneous considerations may account for these unexpected results. First, none of the stimuli photographs was overtly sexual in nature (contrast Rosenzweig, 1942; and Bullock, 1959). Second, participants were not asked to think about anything in particular regarding the photographs, and were instructed merely to look at the pictures at their leisure (contrast Zamansky, 1956).

In another study, Amoroso, Brown, Pruesse, Ware, and Pilkey (1970) found that viewing time for slides increased as a positive linear function of ratings of degree of pornographic content. In a corroborating study aimed on defining "pornography" for future research, Ware, Brown, Amoroso, Pilkey, and Pruesse (1972) measured the viewing time for 40 male college students of 15 slides representing a wide range of sexual behaviors. Participants were asked to rate each slide according to 12 seven-point bipolar semantic differential scales comprised of three factors: evaluative (pleasant – unpleasant, beneficial – harmful, clean – dirty, good – bad); activity (stimulating – nonstimulating, active – passive, hot – cold); potency (heavy – light, strong – weak, serious – humorous); unusual – usual; and an overall pornography rating. Each participant controlled the slide projector and was allowed to view each slide as long as he wished before rating it. The viewing time for each slide was covertly recorded. Statistical analysis showed that the first two semantic differential factors (evaluative and activity) accounted for 94% of the variance in viewing time. In addition, looking time was determined primarily by the activity rating, such that the more stimulating the participant rated the slide, the longer he viewed it. Viewing time of each slide was only moderately related to its evaluative rating. These 40 participants concluded that very "pornographic" material was bad, unpleasant, dirty, harmful, active, hot, and stimulating.

The remaining review of literature in this section is less sequential and more topical. While research into assessment of general sexual interest using sustained visual attention progressed, specific pockets of study regarding external influences on viewing time appeared in the literature during the 1970s and continued through the turn of the century. Research during the 1990s and 2000s also included studies using more precise and accurate viewing time measures aimed at examining both heterosexual and homosexual sexual interest and sexual deviance.

External influences on sustained visual attention. As with many studies involving human subjects, it is possible that external influences may alter or affect a person's behavior.

Audience observation. It is plausible that a person might behave differently in the presence of an audience when asked to complete a task related to a sensitive subject such as sexual interest. In a 1973 study, Brown, Amoroso, Ware, Pruesse, and Pilkey used the same participants and data set as the previous study to measure the effects of an audience on viewing time of sexually explicit stimuli. As the authors expected, "looking time was considerably shorter when the subjects were observed by an audience (mean = 10.43 seconds) than when they viewed the slides alone (mean = 21.16 seconds). This difference is statistically significant (F = 15.88, df = 1/36, $p < .01$)" (p. 131). These results are similar to Martin's (1964) findings that 20 college-age males took longer to sort pictures of nude females under permissive conditions than under inhibitory conditions. From these results in can be assumed that the presence of an audience is at least inhibitory to viewing times, if not threatening to the entire procedure.

Interestingly, an audience study conducted by Saunders and Naus in 1993 yielded no significant effect whether subjects viewed sexually explicit stimuli with a male partner, with a female partner, or alone. Notably, however, the audience conditions in this study involved 80 participants viewing short sex videos alone or with a partner of either gender who was also an active participant in the study. Participants were instructed to rate the videos viewed on several characteristics. Due to each participant's active involvement in the rating activities, it is possible that participants felt neither inhibited nor threatened by a person performing the same activity that he or she was performing. While the reasons may be many for achieving no significant effect for audience condition, this study may have differed from other audience studies in that the "audience" member viewing partners were conceptualized as more actively similar to

participants than as someone who is specifically watching the subject view sexually explicit stimuli.

Personality factors. There are several examples in the literature that suggest that certain personality variables may affect a participant's viewing time of sexual stimuli. In 1976, Love, Sloan, and Schmidt divided 35 male college students into three groups based on their scores on a measure of sex guilt (high, moderate, and low). Sex guilt, as defined for the study and elsewhere, is a "generalized expectancy for self-mediated punishment for violating or for anticipating violating standards of proper sexual conduct" (p. 624; Mosher, 1966). Researchers unobtrusively recorded the amount of time participants spent viewing and rating photographic slides of varying erotic content. Overall, the three groups did not differ significantly in their average viewing time of the slides; but, as the authors hypothesized, for the low sex guilt group, viewing time of specific slides increased as a function of increasing pornographic content (positive linear relationship). For the high sex guilt group, however, there was no significant increase in viewing time. Interestingly, the moderate sex guilt group exhibited a curvilinear viewing pattern. Thus, it appears that for certain individuals (the "priggish"; Love et al., p. 624), perceptions of impropriety regarding such a sensitive subject as sex may result in atypical viewing patterns. In contrast, for persons unconcerned with "self-mediated punishment" regarding social sexual propriety (the "profligate"; Love et al., p. 624), it is likely that viewing times will reflect actual sexual interest, uninhibited by social pressures. As for the moderate sex guilt group (the "prudent"; Love et al., p. 624), the increasing viewing time appears to indicate increasing sexual interest as explicitness increases to a certain degree, at which point it appears that concern for proper sexual conduct overrides viewing behaviors in favor of compliance with social propriety. This curvilinear relationship between stimuli sexual explicitness and viewing time was hypothesized by Brown et al. (1973), but not achieved until Love et al.'s 1976 study.

Other researchers (Lang, Searles, Lauerman, & Adesso, 1980) used similar definitions and division levels of sex guilt to determine if psychological aspects of individual differences can mediate expectancy effects in research on alcohol and social behavior. Lang and colleagues discovered that study participants who thought they had consumed an alcoholic beverage, whether their drink actually contained any alcohol or not, reported greater sexual arousal in response to varying erotic images. In all conditions except the high sex guilt/expect alcohol groups, viewing times increased as a positive linear function of pornography ratings of the stimulus slides.

Kirschner (1976) performed a similar experiment measuring the viewing time of erotic magazines by subjects in a waiting room. His study showed that when another subject also read an erotic magazine (thereby implying approval), subjects low in need for approval showed much longer viewing times than when another subject read a nonerotic magazine (tacit disapproval). Participants with high need for approval were not affected by the actions of other subjects, but generally avoided the erotic stimuli when presented with a choice of magazines in the presence of others. These results are consistent with earlier studies evidencing longer viewing times for more explicit stimuli, particularly of the preferred gender (Amoroso, et al, 1970; Brown, et al., 1973; Martin, 1964; Rosenzweig, 1942; Ware, et al., 1972; Zamansky, 1956). It also seems logical that sex guilt may be related to some of the inhibitory effects described in audience studies (Brown, et al., 1973; Martin, 1964).

A 1995 study by Strassberg and Lowe addressed several personality factors that may create volunteer bias in sexuality research. These researchers hypothesized that volunteers for studies in human sexuality were significantly different from nonvolunteers on several dimensions. Strassberg and Lowe hypothesized correctly and reported evidence that volunteers endorsed a more positive attitude toward sexuality, less sexual guilt, and more sexual experience

than their nonvolunteering peers. Wolchik, Braver, and Jensen (1985) reported evidence that volunteers for participation in sexual psychophysiology experiments are typically more sexually experienced, less concerned about their performance, and have been exposed to more erotica than nonvolunteers. This information has sobering implications for the generalizability of findings for research of sexual topics since Strassberg and Lowe (1995) and Suschinsky, Lalumière, and Chivers (2009) suggest that any study involving subjects who participate on a voluntary basis will, essentially, be using a biased sample that is more sexually open, sexually experienced, and less concerned about their performance than the general population. It is important to remember, however, that due to generalizability restrictions of the Strassberg and Lowe study sample itself, it is unclear how their findings will affect interpretation of the results of concurrent or subsequent studies of sexual issues. Indirect evidence from Chivers and colleagues (2004) is encouraging, however, in its conclusion that while sexual experience may differ, sexual attraction patterns of volunteers and nonvolunteers are likely not different.

Sexual interest and viewing time. In accordance with the literature of the day and his own theories, Singer proposed a "trichotomy of sexual arousal" (1984, p. 232) involving three component responses: the aesthetic response, the approach response, and the genital response. Singer explains the first component:

> An aesthetic response is conceived as an hedonic feeling in response to a sexual stimulus, as, for instance, the sight of an attractive face or figure or a pleasingly stimulating voice. A person displaying such a response might make efforts to keep the object in view by means of eye movements or head turning. (p. 232-233)

Singer also states that the aesthetic response is less susceptible than the approach or genital responses to the effects of classical conditioning. Such a claim may contribute to improved face

validity of measures based on the aesthetic response of arousal than on the other two responses. Viewing time instruments may be said to measure aspects of this aesthetic response.

Quinsey, Rice, Harris, and Reid (1993) studied the relationship between viewing time and ratings of sexual attractiveness regarding male and female preferences of age and gender of sexual objects. Stimuli consisted of 31 slides, half showing a nude female and half showing a nude male from one of four different age categories: infant, child, pubescent, or adult. While all of the photos depicted persons without clothing, none of the models appeared in flirtation poses, therefore none was considered to be sexually explicit. Participants were asked to rate the model in each slide on several measures of his and her physical and sexual attractiveness. Results of this study were mixed: surprisingly, there was only slight variation in viewing times across the gender and age conditions, with the important exception that for both males and females, slides of "both adults and pubescents of the preferred gender were viewed for longer than all categories of the nonpreferred gender ($p < .05$ and $p < .001$, respectively)" (p. 158). While these results are statistically significant, the authors concluded that when sexual explicitness was controlled, viewing time decreased in variability and usefulness and was thus found to be a weaker measure of sexual preference than expected.

Wright and Adams (1994) conducted a study to investigate whether subjects would be distracted by preferred sex stimuli due to an aesthetic response (as described by Singer, 1984), as indicated by longer latencies and/or more errors on a choice reaction time test. This study involved 20 subjects in each of four groups (heterosexual males, heterosexual females, homosexual males, and homosexual females). Participants completed a timed memory task while controlling their viewing time of slides depicting nude male and female images and neutral images. Controlling for age differences between the groups of participants, analysis of covariance results of the choice reaction time test indicate that sexual orientation did affect

cognitive performance: Each of the four groups demonstrated significantly longer viewing time latencies on the preferred sex slides. Santtila and colleagues (2009) corroborated these results in a small-*N* study using a choice reaction time task and sexually explicit and non-explicit images and males and females. According to the result of both of these studies, visual attention discriminated between preferred and nonpreferred sexual objects in nonpedophilic heterosexual and homosexual adults.

In the discussion section of their article, Wright and Adams (1994) proposed that "if this procedure is replicable with similar populations and can be demonstrated to be a reliable and valid measure of normal sexual preference, then populations with deviant sexual arousal (i.e., pedophiles) may be able to be detected utilizing slides that have valence for them (i.e., pre-pubescent children)" (p. 230). Such questions of generalizability to different populations were concurrently being explored. Several studies using similar methods and stimuli were to follow.

For example, in 1996, Harris, Rice, Quinsey, and Chaplin studied the phallometric and viewing time responses of 26 child molesters and 25 heterosexual men to 70 photographic slides forming seven stimuli categories: nude male and female children, nude male and female pubescents, nude male and female adults, and neutral landscapes. Participants were asked to rate the sexual attractiveness of the persons pictured in the slides, and the viewing time of each slide was covertly measured. For the group of nonoffenders, penile tumescence, ratings of sexual attractiveness, and length of sustained visual attention were concordant across stimulus categories. These participants showed greatest physiological arousal, highest report of perceived attractiveness, and longest viewing times for slides of nude adult females with decreasing arousal, ratings, and attention to pubescent and child females and males of any age.

In contrast, the data collected from the child molester participants was discordant among the three variables across the seven stimulus categories. For example, the correlations between

ratings and viewing time were 0.46, $p < 0.001$ for the offenders, in contrast to the nonoffenders' 0.91, $p < 0.001$ for the nonoffenders (these two correlations also differed significantly at the $p < 0.01$ level). While the offenders' ratings of attractiveness did not discriminate this group from the nonoffenders, child molesters did look at slides of children relatively longer than they looked at slides of adults. Overall, however, the offenders' average viewing time ($M = 1.87$ sec, $SD = 0.88$) of all slides was significantly less than the nonoffenders' viewing times ($M = 3.25$ sec, $SD = 3.04$) and showed a more restricted range and low baseline. According to these researchers, the offenders' restricted range of viewing times across all stimulus categories significantly differentiated the offender group from the nonoffender group. Harris et al. (1996) suggested that the offenders' shorter viewing times might indicate a defensiveness, or deliberate inhibition of viewing time across stimuli categories. The child molesters' pattern of viewing times may be similar to the restricted, flat viewing time pattern of increasingly pornographic stimuli of the "high sex guilt" group of participants in the Love et al. (1976) study described earlier in this review. Results of this 1996 study by Harris and his colleagues seem to support the use of viewing time measures when patterns of visual latency can be examined.

Quinsey, Ketsetzis, Earls, and Karamanoukian (1996) tested four predictions regarding correlations between viewing time and ratings of sexual attractiveness and phallometrically measured age and gender preferences. Using a sample of 24 heterosexual females and 58 heterosexual males in two separate studies, the data analysis yielded statistically significant support for three of the four predictions: that male participants would look longer at pictures of their sexually preferred age and gender stimuli than would female participants; that male and female participants would view slides of adults of the opposite sex longer than they would view slides of prepubescent children of both sexes; and that male participants would view slides of pubescent females longer than female participants would view slides of pubescent males.

Separate two-by-four (sex of model by model age category) analyses of variance on each of the three dependent measures indicated that viewing time and sexual attractiveness ratings tended to mirror phallometric data. The fourth prediction—that male participants would view images of adult females longer than female participants would view images of adult males—was supported by this study, but the difference was not statistically significant (F (1, 46) = 1.40 ns). Male participants did, however, look longer than female participants at images of pubescents of the preferred sex (M = 6.44 sec, SD = 6.22 for males; M = 3.42 sec, SD = 3.20 for females; $p < .05$, one-tailed). Overall, this study provided strong support for the idea that ratings of sexual attractiveness are positively correlated with viewing times of both males and females. Phallometric data from the male participants was also positively correlated with both ratings and viewing times.

Even though the 1996 studies by Quinsey and colleagues provide evidence in support of covertly measured viewing time as a measure of sexual preference, the researchers note several characteristics of viewing time measures that may lead to variability in the relationship between viewing time and ratings of sexual attractiveness and phallometric data in other studies. Quinsey et al. suggest that ratings of sexual attractiveness focus participants' attention on the sexual attributes of stimulus images, and the use of phallometry clearly defines the subjects' tasks as sexual in nature. This is not necessarily true when using viewing time. Specifically, these researchers explain,

> Unobtrusively measured viewing time, however, neither focuses subjects' attention on the sexual aspects of the stimuli nor defines the situation as sexual. Viewing time is also unlike penile responding in that it also serves nonsexual functions, such as the general visual appraisal of others. Viewing time, therefore, is more prone to be influenced by

particular procedural and instructional variables than either of these other two measures. (Quinsey, et al., 1996, p. 352)

Such caveats and limitations are important to note when examining the reliability and validity of viewing time as a measure of sexual interest. According to the results of these researchers, viewing time measures—and their accompanying procedural and instructional practices—may need to become more sensitive and specific before they are appropriate for clinical application, although they may already be accurate and precise enough for group research. These authors also suggested that viewing time measures might be advantageous over plethysmography and self-rating methods for studies of sexual interest in children and adolescents (Quinsey et al., 1996).

In the twenty-first century, a study by Gress (2005) compared a viewing time measure with the Sexual Deviance Card Sort developed by Laws and colleagues (2000) and participants' past sexual behavior. Importantly, Gress (2005) found that viewing time of nude and clothed computer-modified images of persons aged 5 years, 9 years, 13 years, and adults produced consistently accurate age and gender sexual preference classifications when compared to past behavior. This research was conducted with adult males who had committed a contact sexual offense ($N = 26$). Later, Israel and Strassberg (2009) found evidence that viewing time is a good measure of categorical sexual interest, but a poor measure of within-category sexual interest for heterosexual men and women. In other words, while the results of their study replicated the Wright and Adams findings that heterosexual men and women viewed opposite sex pictures significantly longer than same sex pictures, Israel and Strassberg discovered that viewing time was relatively insensitive to differences within a given category, such as heterosexual women's evaluations of one man versus other men. This finding is consistent with social cognition theory in that men and women are quick to determine if the stimulus image fits a sexual preference category—reflecting between-category decisions such as recognizing the image as either male or

female, or fitting a desirable age category—but take more time to determine the degree to which the image is sexually attractive. In short, the viewing time measure in the Israel and Strassberg (2009) study yielded longer viewing times for images in sexually preferred categories, but the correlations between the degree of attractiveness for the preferred categories (those viewed longer) did not vary within the preferred categories.

As part of her argument for the necessity of creating valid and reliable measures of sexual interest, Gress stressed the importance of developing an assessment tool for discovering sexually deviant preferences, especially for pedophilia. Results from such assessment tools might be used with sexual offenders to determine treatment needs, to identify high-risk situations, and to predict the possible rate of recidivism (Fischer, 2000; Gress, 2005; Marshall, 1996).

Viewing Time Measures of Sexual Interest

Measurement of sexual interest using viewing time is based on the premise that under certain circumstances people spend more time looking at images that are of personal interest to them than they do viewing images that are not (Amoroso et al., 1970; Crosby, 2007; Gress, 2005; Harris et al., 1996; Israel & Strassberg, 2009; Love et al., 1976; Macrae & Bodenhausen, 2000; Macrae & Bodenhausen, 2001; Quinsey et al., 1993; Quinsey et al., 1996; Rosenzweig, 1942; Santtila et al, 2009; Singer, 1984; Worling, 2006; Wright & Adams, 1994; Zamansky, 1956) Currently, two instruments use viewing time as an assessment of sexual interest: The Abel Assessment for Sexual Interest (AASI) and the Affinity. This study uses the Affinity 2.5, the most recent version of Affinity; however, a brief review of the AASI is theoretically and psychometrically relevant.

Abel Assessment for Sexual Interest (AASI). The AASI is a covert measure of sexual interest as indicated by sustained attention to visual stimuli (Abel, Huffman, Warberg, & Holland, 1998). The assessment tool consists of a paper-and-pencil questionnaire and a 160-

image computer slide show including photographs of fully-clothed men and women of varying ages (Abel, Lawry, Karlstrom, Osborn, & Gillespie, 1994). Following the completion of the questionnaire, a participant views the slides and reports his or her sexual interest in the presented images (Abel, 1996). The participant's viewing time is surreptitiously calculated as he or she views each slide. The calculated viewing time of each image is reported in ipsative form as *z*-scores (Abel, 1996; Fischer, 2000; Krueger, Bradford, & Glancy, 1998), or standardized, unit-free distances from the mean (Warner, 2008). *Z*-score transformation treats the various responses of each subject as a distribution of scores, calculates the mean and standard deviation of these scores, and transforms each response into a standard normal deviate of the distribution (Barbaree & Mewhort, 1994). Theoretically, these AASI *z*-scores show the relative strength of interests that an individual demonstrates toward each of the different image categories. Raw data from the test administrations are sent to Abel Screening in Atlanta, Georgia, and the computed results are returned electronically to the test administrators (Krueger et al., 1998).

Strengths of the AASI include its standardized administration procedures, slide sets, and testing format (Abel, et al., 1998; Kaufman, Rogers, & Daleiden, 1998; Smith & Fischer, 1999); its availability for use with males and females of a wide age range; its nonpornographic, sexually nonexplicit stimulus images; and its noninvasive focus on sexual interest, as opposed to plethysmography's physiological focus on sexual arousal. The AASI is also built on reasonable visual attention theory, the technology is readily available to researchers, and the structure for further research is in place (Fischer & Smith, 1999). The AASI is the most popular tool for assessing sexual interest using viewing time. Even so, some researchers suggest the need for its further development and refinement (Fischer & Smith, 1999; Smith & Fischer, 1999).

Despite the AASI's many strengths, several studies using samples of known sex offenders question its adequacy and highlight its limitations. In 1998 Kaufman, Rogers, and

Daleiden presented a study using the AASI with a sample of juvenile sex offenders. These authors examined the test-retest reliability of scores on the AASI and found such reliability to be lower than expected. Kaufman and colleagues suggested the need for more research regarding the temporal stability of AASI scores for adolescents. Shortly thereafter, Gray (1999) used a sample of confirmed pedophiles to study the effect of dissimulation on AASI results. He found that among a group of dissimulators, the AASI was able to identify only 36% of the pedophiles. Gray underscored the need to consider the effects of dissimulation on AASI scores when interpreting assessment results, and suggested that researchers re-evaluate how the AASI is used. In a later study, Letourneau (2002) compared trimmed and untrimmed data from the AASI to data obtained via penile plethysmography from a sample of incarcerated sex offenders. This comparison was notable because the AASI typically reports only the trimmed data, with outliers removed according to a confidential formula (Fischer & Smith, 1999). Abel's use of trimmed data is problematic because it disallows computation of coefficients alpha, and thus prohibits evaluation of internal consistency. Letourneau concluded that both the AASI and plethysmography have strengths and weaknesses, and both need improvement for use in the assessment of sex offenders.

In associated articles, Fisher and Smith (1999) and Smith and Fischer (1999) questioned the competence of the AASI for use as currently marketed as a sexual deviance screening tool for adolescents. These authors detailed multiple concerns about the reliability and validity of the AASI and reported results of several studies conducted by various researchers yielding substandard (coefficients < .80; Anastasi, 1988) reliability coefficients for the data gathered using this measure. Specifically, Smith and Fischer (1999) reported that analysis of test-retest data did not support the reliability of the measure for use with adolescents, screening validity data showed that the ability of the instrument "to discriminate adolescent offenders from

nonoffenders was not significantly better than chance" (p. 214), and diagnostic validity data showed that the ability of the tool to identify specific deviant attractions (such as sexual attraction to children, for example) within the known adolescent perpetrator group was poor. According to these researchers' detailed analysis and review of the literature, the test maker's claims of the reliability and validity of the AASI to screen or diagnose adolescent perpetrators (Abel et al, 1994; Abel, 1996; Abel, 1997; Abel et al., 1998), are questionable at best and could benefit from further study.

Fischer and Smith (1999) and Fischer (2000) also questioned the psychometric adequacy of the AASI and the data reported by the test makers. Raw data are unattainable, and AASI data are reported as ipsative *z*-scores, or indications of intraindividual variation. This provided ipsative information is incomplete, however, as Fischer and Smith explain:

> Interpretation of intraindividual variation is enhanced if one is aware of the underlying mean and standard deviation from which the ipsative scale was created. In the absence of such information, ipsative scales are analogous to ordinal scales. In such cases, interpretations are limited to conclusions that a subject possesses more of one attribute than another attribute, but it is impossible to state how much more he [*sic*] possesses of one than another or when differences between attributes are significant. (1999, pp. 196-197).

Specifically, because the AASI raw score means and standard deviation means for the viewing time results are not provided, misinterpretation of the data as norm-referenced interval scores is likely and inadvertently misleading (Fischer & Smith, 1999). With a lack of a normative reference group, Fischer (2000) also explains that interpreters of the reported data are unable to compare any given participant to another because there is no normative baseline or interval in which to ground the interpretation of the scores.

There are other concerns about the participant data reported by the test creators. According to Ewing (2006), the ruling in a recent court case specifically excluded AASI test results, concluding that the scores from the assessment did not meet the standards for admissibility as scientific evidence. In *Commonwealth of Massachusetts v. Gerard Ready* (2005), the court criticized the lack of published studies on the AASI, the tool's high error rate, the requirement that all raw data be sent to the test creator to be processed, and the fact that underlying formulas and modifications to the data (removing outliers, etc.) are not made known. The court stated, "For all we know, they and their components could be mathematically based, founded upon indisputable empirical research, or simply the magic of young Harry Potter's mixing potions at Hogwarts School of Witchcraft and Wizardry" (Ewing, 2006, p. 61). Apparently the court has little faith in the data reported by the AASI, particularly because of their purposely-shrouded method of derivation.

Affinity 2.5. As described previously, the theory underlying viewing time measures is that when asked to rate the sexual attractiveness of a particular model, individuals look longer at, and take longer to categorize, images they deem sexually interesting relative to the amount of time spent looking at and categorizing images that they deem sexually uninteresting (Amoroso et al., 1970; Crosby, 2007; Gress, 2005; Harmon, 2006; Harris et al., 1996; Israel & Strassberg, 2009; Quinsey et al., 1993; Quinsey et al., 1996; Rosenzweig, 1942; Santtila et al., 2009; Worling, 2006; Wright & Adams, 1994; Zamansky, 1956). Affinity 2.5 is a computer-based assessment instrument developed to measure such sexual interest as indicated by viewing time (Glasgow & Fischer, 2006a). Affinity 2.5 creates individual profiles of relative sexual interest by age and gender. These profiles are generated as the participant completes two tasks: a ranking task, and a rating task. The ranking task presents a series of eight line drawings representing males and females of various ages. The participant is asked to rank the drawings from most

sexually attractive to most sexually unattractive. Upon completion of this preliminary ranking, the participant is directed to rate on a sliding 15-point scale the sexual attractiveness or unattractiveness of 80 color images of fully-clothed males and females in different developmental stages from small child to adult. The participant's viewing time of each image is covertly measured during this rating task.

The intention of the Affinity is not to assess sexual arousal, but expressed sexual interest (Glasgow, Osborne, & Croxen, 2003). Because of this distinction in what is actually being measured, the Affinity does not need to use pornographic or nude images of real or composite persons. This is an important improvement over previous studies and assessment instruments that confronted some understandable ethical concerns regarding the appropriateness of showing to study participants images of people wearing little or no clothing. Laws and Gress provide a summary of the literature using viewing time assessment of sexual interest utilizing computer-constructed images of stimuli subjects of different ages and stages of physical and sexual development (Gress, 2001; Laws et al., 2000; Laws & Gress, 2003; The Pacific Psychological Assessment Corporation (PPAC), 2004; Vanstone & Laws, 1998). Laws and Gress (2004) have developed two image sets of computer-constructed person-like images of males and females. These image sets have yet to be used in published, formal research studies, but their invention and validation has important ramifications for studies of viewing time and sexual interest since they provide an alternative to photographs of real people. The image developers maintain that using computer-constructed composite images of people, as opposed to photographs of real people, avoids some ethical dilemmas present when showing to study participants images of nude or partially-clad models. Such claims are interesting, but arguable. The Affinity addresses such arguments by using only pictures of fully-clothed males and females of various ages.

There has been some research conducted on the Affinity instruments as measures of sexual interest. In 2006 Worling used the Affinity 1.0 self-report ratings of sexual attractiveness, the Affinity viewing time measure, and another self-report sexual arousal graphing procedure with 78 adolescent males who were known sex offenders. Worling (2006) reported that the pattern of responses to all three assessment techniques was remarkably similar. In addition, the Affinity viewing time approach significantly differentiated those adolescents who assaulted male children from those who assaulted other individuals. In 2007, Cloyd conducted a concordance study of plethysmography data and Affinity 2.0 data from 96 known adult male sex offenders. She discovered the data from these two different assessment tools to be significantly correlated--a fact which supported the validity of Affinity 2.0 as a measure of sexual interest.

Affinity 2.5 resolves some of the problems discovered in the AASI by reporting raw score results that have not been transformed by standardization. Because Affinity 2.5 reports all raw scores for each task, comparisons between participants are possible. While the Affinity data reported are ipsative (as in the AASI), rather than normative, the data are in raw form and have not been standardized. Glasgow et al. (2003) explain that data transformations of ratings and viewing latencies are "contraindicated here because the ratings are far from normally distributed, and it is also often the case that latency measures are significantly skewed" (p. 99). These researchers echo the data transparency standards proposed by Krueger et al. (1998) that any transformations must be both conservative and not misleading. The assessment results produced by Affinity 2.5 are a figuratively conservative step toward reporting clarity.

Factors Related to Psychometric Adequacy

Certain principles of psychometric and technical adequacy can be used to evaluate any device. Examination of the nature of an instrument's data and its reliability, validity, and norms is imperative.

Ipsative and normative scores. A remaining drawback for both the AASI and Affinity 2.5, is their use of ipsatized scores. Mathematically, ipsative scores always sum to a constant (Clemans, 1956). In technical terms, the AASI converts raw viewing times into ipsatized *z*-scores overlayed by an arbitrary rule of thirds (Abel et al., 1998; Smith & Fischer, 1999). Raw data from the Affinity 2.5 is ipsatized into mean rank scores per category (Glasgow, 2003). While on the Affinity no other transformations are automatically calculated, there is some distortion in the pattern of the raw data that occurs during ipsatization: It is possible that large differences between scores can be minimized and minor differences can be exaggerated in the conversion process (Brown, 2005; Madsen, 2008). This is particularly problematic if these ipsatized scores are misinterpreted as interval or ratio data, which is incorrect, and comparative conclusions are drawn. In reality, converting raw data to ordinal data tends to confound interpretation because it removes the equal intervals (in this case, seconds or milliseconds of viewing time) between ticks on the scale. Madsen (2008) explains:

> In making all raw scores sum to a constant, any relation to the absolute value of the attribute is lost. The scale of the scores is lost. Intervals are lost. Conclusions about the degree of individual traits cannot be based on ipsative scores. A high ipsative score does not necessarily equate to a high degree of an attribute's actual value. Scores have meaning only in relation to other scores within the ipsative profile. Unfortunately, with the loss of scale, it is equally unclear when differences between scales are significant. One can tell neither the absolute strength of an attribute nor the importance of any differences between attributes. (p. 18)

For the Affinity, the scoring interpretation procedures represent one of the major areas of continuing research.

According to the test creator (Glasgow, 2003), the preferred method of interpretation of data for an earlier version of the Affinity was visual analysis of patterns (usually line graphs) in comparison to the client's self-report. This was an appropriate method of interpretation of ipsative scores. Using this type of ordinal data, it is important to avoid any conceptualization of the information as an indicator of normalcy or deviancy since the only comparisons that can be made are intraindividual, not interindividual (Fischer & Meade, 2010; Fischer & Morgan, 2006; Warner, 2008). Using ipsative scores, assessors are limited in their ability to identify and diagnose deviance because they have no standard against which to compare an individual's assessment results.

In contrast, most widely used psychometric devices are norm-referenced, which allows comparison of the observed scores of one individual on a specific scale to the distribution of scores acquired from a reference group on the same scale. This is a common method of interpretation and allows for conclusions of normalcy regarding the individual being assessed. No such normative group exists for the AASI, but progress has been made in this area with the Affinity. Harmon (2006) and Crosby (2007) studied the viewing time mean response patterns on Affinity 2.0 of females and males with sexual interests reported as exclusively heterosexual and nonpedophilic. These researchers empirically confirmed intuitive predictions about typical nonpedophilic heterosexual response patterns with data showing that females looked significantly longer at images of adult and juvenile males than they looked at images of adult and juvenile females and pre-juvenile and small child persons of either gender (Harmon, 2006). Crosby (2007) found that nonpedophilic, exclusively heterosexual males looked significantly longer at images of adult and juvenile females than they looked at images of adult and juvenile males and pre-juvenile and small child persons of either gender. Worsham (2009) and Boardman (2009) replicated Harmon's (2006) and Crosby's (2007) studies using Affinity 2.5 with very

similar results. The participants in these studies represent the normative samples against which the Affinity response patterns of other nonpedophilic, exclusively heterosexual females and males could be compared.

Research on this norm-referenced approach to interpretation of Affinity scores has continued. Fischer (Fischer, 2004; Fischer & Meade, 2010; Fischer & Morgan, 2006; Glasgow & Fischer, 2006a; Glasgow & Fischer, 2006b) has proposed a chi-square goodness-of-fit method of comparing an individual's observed pattern of ipsative scores to a normative expected pattern of ipsative scores. Fischer and Meade (2010) explain that this chi-square logic "assumes that there is a typical expected pattern and that deviance from that expected pattern may occur in myriad ways. The chi-square residuals identify any number of patterns that are deviant from expectation" (p. 21). Such comparisons promote understanding about normalcy and deviance and facilitate decisions that can be used for screening and diagnosis. The overall chi-square value can represent the goodness-of-fit and function as the screening portion of the test; and the chi-square residuals can represent the specific areas of deviance of the observed pattern from the expected pattern and function as the diagnostic portion of the test (Fischer & Meade, 2010). Of course, any conclusions about an individual's pattern of deviance should correctly be stated in terms of the characteristics of the reference group pattern to which the individual pattern was compared.

Fischer's approach has been applied to data from several studies to determine if it can successfully identify and diagnose known sex offenders (Fischer, 2006a; Fischer, 2006b; Fischer, Byrne, & Glasgow, 2007; Fischer & Morgan, 2006). These studies involve data collected using early, now obsolete versions of the Affinity, but generally yield high concordance rates between the Affinity data and offender profiles including ages and genders of the known victims. These studies are likely to advance the utility and efficacy of viewing time measures of sexual interest in the screening and diagnosis of sex offenders.

Test-retest reliability as an estimate of temporal stability. As Fischer (2000) explains, exploring an instrument's test-retest reliability is crucial to the development and validation of any assessment tool that purports to measure a construct considered to be relatively constant across time. Sexual interest is one such stable construct. Reliability is defined as "consistency of measurement results" (Warner, 2008, p. 830). A reliable instrument measuring a stable construct would be expected to produce highly similar results under two different test administrations. The reliability of scores obtained from a testing procedure can never be determined exactly, but it can be estimated. To assess reliability, a researcher needs to gather at least two sets of data for the same construct—in this case, Affinity viewing time data regarding sexual interest from the same participants at two different times—and calculate an appropriate statistic to assess the consistency, or stability, of the scores across time.

Harmon (2006) calculated estimates of test-retest reliability, or temporal stability, for Affinity 2.0 scores using the Pearson Product Moment Correlation Coefficient (PPMCC) statistic for the category means and medians at time one (T1) and time two (T2), dates at least two weeks apart. Unfortunately, data about individual categories on the Affinity (such as the PPMCC statistics) are only informative in the context of the overall pattern in which they exist. Crosby (2007) also calculated PPMCC statistics as estimates of the temporal stability of Affinity 2.0 scores, and added Spearman's Rank Correlation Coefficients and chi-square statistics. Worsham (2009) and Boardman (2009) calculated temporal stability estimates for Harmon's (2006) and Crosby's (2007) Affinity 2.0 data in terms of the overall patterns obtained at T1 and T2 using Fischer's chi-square goodness-of-fit method mentioned previously: Viewing time data from T1 served as the expected pattern, and viewing time data from T2 were designated as the observed data. According to the chi-square logic, the greater the fit between the patterns at T1 and T2, the greater the reliability of the Affinity instrument.

Worsham (2009) found that 86% of the participants in Harmon's (2006) study (all nonpedophilic, exclusively heterosexual females) had score patterns evidencing the temporal stability of Affinity 2.0, while 14% of the participants did not. Similarly, Boardman's (2009) analysis of the Crosby (2007) data showed that 77% of participants (nonpedophilic, exclusively heterosexual males) showed evidence of temporal stability of their Affinity 2.0 score patterns, while 23% did not. These temporal stability estimates are positive indications of Affinity 2.0's reliability, but they are clearly less than perfect. According to these analyses, even though 86% and 77% of participants' patterns of scores were stable across time, respectively, 14% of the females sampled, and 23% of the males sampled recorded response patterns that were significantly different from each other at T1 and T2. Worsham and Boardman admit that while these numbers represent far less than a majority, they do indicate a need for caution when interpreting the Affinity 2.0 scores.

Affinity 2.5 is the newest version of this viewing time measure of sexual interest. Affinity 2.5 includes 42.9% more stimulus images than Affinity 2.0. In general, increasing the number of good-quality items in an instrument tends to increase the reliability of the resulting scores. As the primary portions of their dissertations, Worsham (2009) and Boardman (2009) established an expected normative pattern of responses necessary in order to effectively employ Fischer's chi-square method of data analysis (Fischer, 2004; Fischer & Meade, 2010; Fischer & Morgan, 2006; Glasgow & Fischer, 2006a; Glasgow & Fischer, 2006b). The current study will take the next step in exploring the validity of Affinity 2.5 as an effective and useful measure of sexual interest by studying the instrument's temporal stability.

Method

The following participants, procedures, measures, and data analyses were employed in the completion of this study.

Participants

Participants included exclusively heterosexual, nonpedophilic male and female college students, age 18 or older, recruited from Brigham Young University (BYU) in Provo, Utah. Because this study specifically sought responses of exclusively heterosexual nonpedophilic adult males and females, researchers included in the data analysis only information from students identifying themselves as exclusively heterosexual and claiming no history of pedophilic interest. All students who participated in this research received extra credit in their referring class as allowed by their instructors.

In order to determine a practically and statistically appropriate sample size for this study, the researcher used the sampling distribution analysis method described by Worsham (2009) and Boardman (2009). These researchers used Harmon's (2006) and Crosby's (2007) Affinity 2.0 data to determine the ideal number of participants for their studies. The central limit theorem states that as sample sizes increase, the sample means more accurately estimate the true mean of the population (Howell, 2002); so, in general, larger sample sizes yield more accurate representations of the population. It is also true that as sample sizes increase, the variance of the sampling distribution decreases. Variance is defined as the average of the squared deviations of the observations in a data set about their group mean. If a distribution has a relatively smaller variance, the observations in that distribution will tend to be clustered closely about their mean. Reflecting the central limit theorem, there is a sample size at which the number of participants

provides the most useful amount of central tendency while still providing important variance. It is that sample size that Worsham and Boardman sought to discover.

Using the Affinity 2.0 data, sampling distributions were analyzed to calculate and graph mean response curves and mean variances across the eight different categories of images. Worsham (2009) and Boardman (2009) visually examined these data in order to determine the point at which the patterns stabilized. They concluded that the mean proportions of total time spent in each category began to stabilize with about 50 participants. The standard deviations of the proportions of total time spent viewing each image category also stabilized and grew smaller with a sample of about 50 participants. Sample sizes larger than 50 contributed only a minimal amount of stability. Although the current study is an examination of the temporal stability of Affinity 2.5, and not Affinity 2.0, the researcher determined that sample sizes of at least 50 exclusively heterosexual nonpedophilic females and 50 exclusively heterosexual nonpedophilic males were adequate to meet the practical and statistical demands of this study.

Procedures

The researchers believe that it is very important to protect the confidentiality of participant responses in this study. When an individual presented for participation in this project, he or she was assigned a participant number attached to a single master list matching participant names and numbers. The master list was kept in a locked file to be accessed only by members of the research team. No names were recorded in the computer database or on the questionnaires.

Participants were instructed to review and sign informed consent documentation (Appendix A) detailing the nature of the study for which they were volunteering. After signing the consent form, participants were directed to a private room designated for the purpose of this research, and received instructions in order to complete the 15- to 20-minute computerized assessment of sexual interest. Once the researcher had given instructions and answered any

questions, he or she exited the room until the participant had finished. Upon completion of the assessment, the participant completed a questionnaire designed to gather information on demographics, social desirability, and sexual interest. Because this study is a test-retest measure of stability of scores across time, following this initial administration of Affinity 2.5, participants were requested to return in no fewer than 14 days to complete Affinity 2.5 for the second time.

Measures

Affinity 2.5. Affinity 2.5 is a computer-based assessment instrument that measures a person's relative sexual interest in images of males and females of four different age categories. The instrument creates a profile for each participant as the person completes a ranking task and a rating task. The original version of the Affinity was designed to assess the sexual interest of males with learning disabilities. The current version of the assessment tool is licensed for use with disabled adult male sex offenders and nondisabled male sex offenders. The test's creator, David Glasgow (2003), has also approved the use of Affinity 2.5 for research and evaluation of adult male nonoffenders, juvenile male offenders, and female offenders.

The ranking task begins with the participant viewing and ranking several prototype images—simple line drawings depicting a person from each of eight categories: adult male, adult female, adolescent male, adolescent female, preadolescent male, preadolescent female, small child male, and small child female. The participant is asked to rank order the line drawings from most sexually attractive to most sexually unattractive, beginning with the drawing depicting the type of person he or she considers to be the most sexually attractive. The participant continues to select attractive drawings until he or she decides that the remaining images are no longer sexually attractive. The participant then ranks the remaining images according to their sexual unattractiveness. The purpose of this prototype ranking procedure is to predict the rank order of each stimulus category when ranked by viewing time and attractiveness in a subsequent task.

Ultimately, the viewing time portion of the Affinity serves as a test of honesty of the participant's self-report ranking of sexual interest (Glasgow, 2003).

The rating task requires the participant to view and rate the sexual attractiveness of several practice images and 80 test images. Each of the eight total male and female prototype categories used in the ranking procedure is represented by 10 photographic images. The participant is instructed to view each image and then rate the image's sexual attractiveness using a 15-point sliding scale ranging from "very sexually attractive" to "very sexually unattractive." As the participant completes this rating procedure, two measures of viewing time are surreptitiously recorded. The first measure of viewing time recorded is On-Task Latency (OTL), which is the time elapsed from the initial presentation of the image on the screen to the time that the participant rates the image. The second measure of viewing time recorded is the Post-Task Latency (PTL), which is the time elapsed from when the participant rates the image to the time the image is changed.

All viewing time measurements are reported in raw score form in seconds and milliseconds and converted to mean ranks (Glasgow, 2003). Images are rank ordered from the longest-viewed to the shortest-viewed and assigned a value from 1 to 80. The rank scores of the images in each category are then averaged to provide a mean rank of images in each category. These mean rank values are ipsatized raw scores that sum to a constant of 317. This allows the computerized graphic representation of each person's profile to be consistent across participants, but the true underlying raw score behavior is lost.

The Demographics, Attitudes, and Sexual Interest Questionnaire (DASIQ). The DASIQ (Appendix B) was designed specifically for Affinity studies and used in previous projects (Boardman, 2009; Crosby, 2007; Harmon, 2006; Worsham, 2009). As indicated by its name, the DASIQ consists of three parts. The first part gathers demographic information about

the participant, including questions regarding age, ethnicity, year in school, and marital status. The second part is a condensed version of the Marlowe-Crowne Social Desirability Scale (M-C SDS) called the M-C 2(10). This shortened version of the original instrument was designed to examine an individual's socially-desirable responding within a limited administration time. As a much shorter, simpler measure than the M-C SDS, the M-C 2(10) has demonstrated only a tolerably mild attendant drop in reliability from the original instrument (Strahan & Gerbasi, 1972). The third part of the DASIQ is a sexual preference inventory adapted from the Kinsey Heterosexual-Homosexual Scale (Kinsey, Pomeroy, & Martin, 1998). This portion of the questionnaire requires the participant to report his or her sexual preference as described by one of seven different categories.

Data Analysis

This project is an examination of the temporal stability of the data gathered using the Affinity 2.5. In contrast to previous studies (Crosby, 2007; Harmon, 2006), the researcher employed a chi-square goodness-of-fit approach only (as opposed to computing other correlation coefficients) in order to determine how well a participant's overall viewing time pattern at T1 matched his or her pattern at T2. The greater the fit between the two patterns, the greater the reliability of the Affinity instrument. T1 responses were used as the expected pattern and compared to the T2 responses, designated as the observed pattern. The chi-square coefficients were calculated by multiplying by a constant (n) the sums of the differences between the T1 expected proportions (π) and the T2 observed proportions (P) according to the following equation:

$$\chi^2 = n_{\bullet} \sum_{j=1}^{J} \frac{(P_j - \pi_j)^2}{\pi_j}$$

An expected normative pattern is necessary in order to effectively employ Fischer's chi-square method (Fischer, 2004; Fischer & Morgan, 2006; Glasgow & Fischer, 2006a; Glasgow & Fischer, 2006b) in the interpretation of Affinity data. In previous studies Boardman (2009) and Worsham (2009) established such normative patterns for nonpedophilic, exclusively heterosexual males and females, respectively.

Results

Eighty-six males and 105 females participated in this study. Only data from participants who completed both administrations of the assessment (T1 and T2) were included in the analysis, leaving 67 males and 90 females. In addition, data from four of the males and six of the females who took the test twice were excluded from analysis. Three of these male participants and four of these female participants endorsed an item on the Kinsey scale other than "exclusively heterosexual with no homosexual interest," and thus did not meet criteria for inclusion in this study. One male and two females were also excluded for invalid test administrations. The adjusted numbers of participants who were included in the final analysis were 63 males and 84 females.

The remaining participant ages ranged from 18 to 51 years for the males, and 18 to 56 years for the females. Table 1 details the reported demographics of the participants included in the data analysis.

Table 1

Participant Demographics

	Males (n = 63)		Females (n = 84)	
	Number	Percentage	Number	Percentage
Ethnicity				
Caucasian	51	81	72	86
Hispanic	8	13	6	7
Asian	2	3	5	6
African American	2	3	1	1
Year in school				
Freshman	27	43	40	47
Sophomore	13	21	17	20
Junior	8	13	16	19
Senior	13	21	10	12
Graduate student	2	3	1	1
Marital Status[a]				
Single	54	86	78	93
Married	9	14	6	7

[a] None of the participants reported his or her marital status as divorced or widowed.

These percentages appear to reflect the ethnic composition of the student population at this university, but likely differ from that of other universities or geographical regions of the United States.

As previously explained, each participant in this study completed Affinity 2.5 twice. Participants were asked to return to complete Affinity 2.5 the second time no less than 14 days following the first administration. The actual mean number of days between T1 and T2 was 15 (for both males and females), with a standard deviation of 5.4 days for the males, and 4.6 days for the females.

Expected Normative Pattern

In order to address problems with interpretation of ipsative data, several studies have been completed to establish norms against which interindividual comparisons can be made. In addition, an expected normative pattern is necessary in order to effectively employ Fischer's chi-square method (Fischer, 2004; Fischer & Morgan, 2006; Glasgow & Fischer, 2006a; Glasgow & Fischer, 2006b) in the analysis of Affinity data. Worsham (2009) established just such a normative pattern for her sample of 63 exclusively heterosexual, nonpedophilic females by calculating the average category proportions of total viewing time for each of the eight categories of images on the Affinity 2.5. Boardman (2009) followed the same procedures for his sample of 50 exclusively heterosexual, nonpedophilic males. The same calculations were performed for the current study in order to determine if the collected data is appropriate for Fischer's chi-square method. Participants in the current study identified themselves as exclusively heterosexual and nonpedophilic and are demographically similar to those who provided the data for Worsham's and Boardman's studies (Crosby, 2007; Harmon, 2006). Data reflected in the characteristic curves from these earlier researchers constitute the normative samples against which the data from the current study are compared. Because the participants in Worsham's and Boardman's

studies only completed the Affinity 2.5 once, we used the current study participants' proportion means from T1. The proportion means and visual patterns of responses for both the females and males in the previous studies and the current studies are detailed in Table 2 and Figure 1.

Table 2

Comparison of Affinity 2.5 Proportion Means

	Mean Proportion of Total Viewing Time by Category							
Study Participants	ADF	JUF	PJF	SCF	ADM	JUM	PJM	SCM
Worsham (2009) Females	.124	.115	.101	102	.189	.164	.113	.092
Current Females (T1)	.126	.109	.100	.110	.187	.165	.107	.095
Boardman (2009) Males	.205	.179	.115	.109	.104	.105	.093	.089
Current Males (T1)	.197	.177	.115	.116	.107	.101	.096	.089

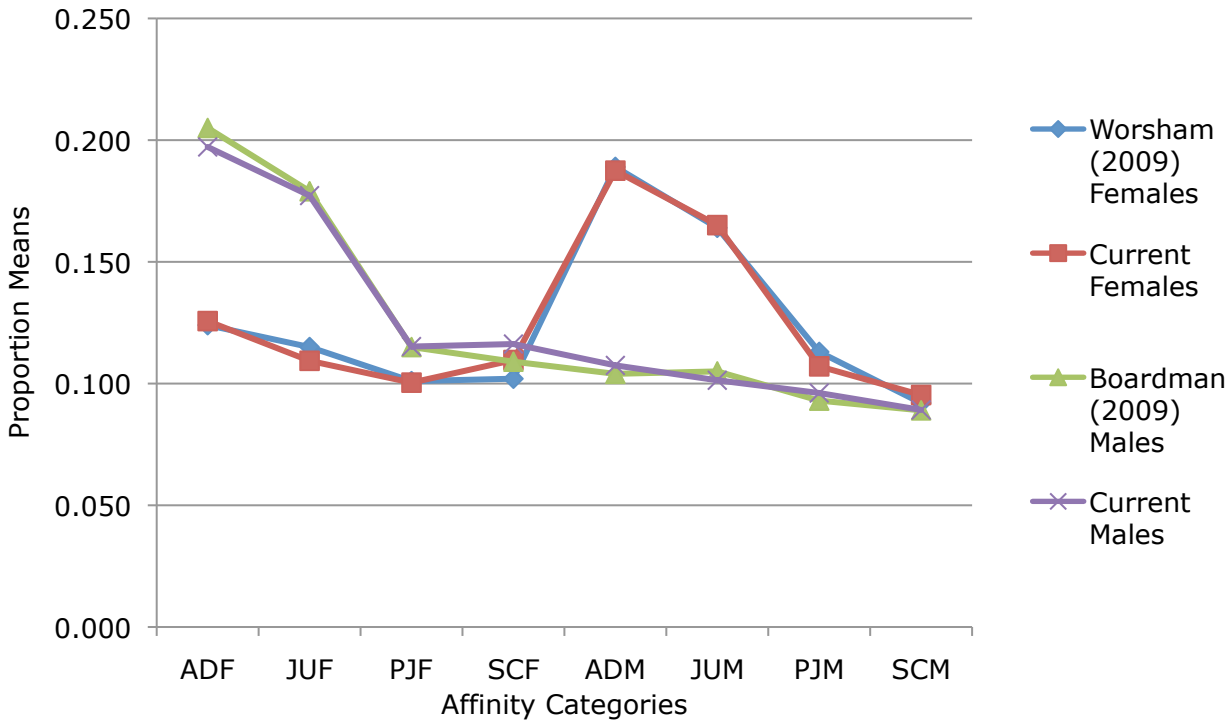

Figure 1. Proportion means for previous and current studies of Affinity 2.5. Comparison of viewing time proportion means for Affinity 2.5 from normative samples of females and males established in Worsham (2009) and Boardman (2009) with viewing time proportion means of the current samples of females and males.

As shown in the table, the proportion means established in this study using Affinity 2.5 are analogous to those established in Worsham's (2009) and Boardman's (2009) studies using Affinity 2.5. In addition, the patterns of proportion means across Affinity 2.5's eight image categories for the current research are visibly similar to the 2009 studies, suggesting that the current data is similar enough to the normative data for this assessment tool that the use of Fischer's chi-square method of analysis is appropriate.

Establishing characteristic curves of responses typical of populations of different types of people is necessary for making statements of normality, abnormality, and deviance. These studies establishing norms for Affinity 2.5 provide visual evidence of a temporally stable, normal curve from which statements of normality and abnormality can be more accurately made. Potential uses of Affinity 2.5 as a screening tool for sexual deviance depend upon these proportion means and normative pattern of responses for this population. Screening cannot occur without such norms, since logic and qualities of the ipsative data do not support it. As research progresses, it will be important to establish local norms for populations with different sexual orientations and other demographic characteristics.

Chi-Square Coefficient Calculation

Chi-square coefficients were calculated by multiplying by a constant (n) the sums of the squared differences between the T1 expected proportions (π) and the T2 observed proportions (P), as described previously. Fischer and Meade (2010) explain that converting the raw observed and expected viewing times into proportions standardizes each subject's viewing time results. These new, standardized values can then be multiplied by any standard factor (n) for which there is a viable rationale. The constants utilized in the current analysis were 115 for the males, and

116 for the females. The rationale for using these values is worthy of explanation. We have stated previously that with more items, we would expect the longer Affinity 2.5 measure to be more stable than the shorter Affinity 2.0 measure. Even so, the chi-square statistic is incredibly sensitive to minor variations among the data such that plans for comparison of chi-square values require measures be taken to ensure some form of consistency across different versions of the measure, among participant groups, or in the calculation of their resulting data. One way to address this need for consistency, and thus facilitate comparison between the two versions of the test, is to use the same constant in all of the chi-square calculations for the females and the same constant in all of the chi-square calculations for the males.

The search for appropriate fair multipliers to be used in the calculations of the chi-square coefficients in the current study led the researchers to Worsham's (2009) analyses of Harmon's (2006) Affinity 2.0 sample of 120 exclusively heterosexual, nonpedophilic females. Worsham used a constant of 116, which represents the median of the total viewing time at T2 for Harmon's female sample. This value was used as the multiplier for calculation of chi-square values for the females in the current sample. A similar number was found in the median of the total viewing time at T2 for Crosby's male sample. In 2009, Boardman analyzed Crosby's (2007) Affinity 2.0 data collected from 77 exclusively heterosexual, nonpedophilic males to examine the temporal stability of their scores on that measure. In his report, Boardman identified 148 as the constant in his chi-square calculations. For consistency, and to facilitate comparison of like chi-square scores, the current researchers recalculated chi-square values for Crosby's Affinity 2.0 data (included in Appendix C) using the median of the total viewing time at T2 for his male sample (115). The researchers also used that value as the multiplier for the current analyses of the data collected from the males in this study.

Affinity 2.5 includes eight categories of stimulus images: Adult Female (ADF), Juvenile Female (JUF), Prejuvenile Female (PJF), Small Child Female (SCF), Adult Male (ADM), Juvenile Male (JUM), Prejuvenile Male (PJM), and Small Child Male (SCM); thus, the degrees of freedom for this analysis is seven. At a significance level of .05, the chi-square critical value is 14.067. Using this critical value, 10 of the 84 female participants had significant chi-square coefficients. Nine of the 63 male participants had significant chi-square coefficients. A significant chi-square value indicates inconsistency, or instability, in the participant's pattern of responses from T1 to T2. Nonsignificant scores were found for 74 of the 84 females and 54 of the 63 males. According to this chi-square approach, 86% of the male participants and 88% of the female participants had scores evidencing the temporal stability of Affinity 2.5. Appendix D contains a list of the chi-square values for the males in the current study, and Appendix E lists the chi-square values for the females.

Table 3 allows for comparison of the temporal stability of the shorter Affinity 2.0 with the longer Affinity 2.5.

Table 3

Percentage of Stable, Nonsignificant Chi-Square Values for Affinity 2.0 and Affinity 2.5

	Males	Females
Affinity 2.0	81%[a]	86%[b]
Affinity 2.5	86%	88%

[a] Value recalculated from Crosby (2007) using 115 as the constant multiplier. See Appendix C.
[b] As reported in Worsham (2009).

It appears from these analyses that the scores on Affinity 2.5 for samples of exclusively heterosexual, nonpedophilic males and females are slightly more stable than the scores on Affinity 2.0 for similar samples of males and females.

As previously detailed, participants in this study were requested to return for the second administration of Affinity 2.5 at least 14 days following their first administration. Because there was variation among the subjects in the number of days between test and retest, the researchers tested whether there was a difference between subjects whose viewing time was stable (nonsignificant chi-square coefficients) and those whose viewing time was unstable (significant chi-square coefficients) in terms of days between testing. Parametric independent samples *t* tests were conducted on all subjects, as well as subgroups (males only and females only) to determine whether there were significant differences in days between testing. No significant differences were found ($t_{all}(145) = -1.002$, $p = .318$; $t_{males}(61) = .535$, $p = .594$; $t_{females}(82) = -.844$, $p = .420$). Because the distributions of days between treatment were skewed, nonparametric Mann-Whitney *U* tests were conducted as well. No significant differences were found ($U_{all} = 1205.00$, $p = .934$; $U_{males} = 211.50$, $p = .407$; $U_{females} = 340.00$, $p = .564$). According to these analyses, the interval between test and retest did not account for subject viewing time stability or instability.

Discussion

As previously detailed, results of this analysis indicate that 86% of responses from the male participants and 88% of responses from the female participants were stable across time. Although these numbers represent respectable levels of temporal stability, and an increase over the temporal stability of the shorter Affinity 2.0, 14% of male participants and 12% of female participants recorded response patterns that were significantly different from each other from T1 to T2. The implications of these *instabilities* seem to deserve some exploration.

Questions and Implications

A series of questions is intriguing: Given that one in seven males and one in eight females are likely to respond to this measure significantly differently from T1 to T2, are people generally unstable in their viewing time of images of individuals they are evaluating for sexual attractiveness? What does instability of responses mean if it occurs for some people but not for others? Could such instability become a diagnostic consideration; and if so, how might it be used? Such questions are theoretical in nature, but have practical implications. Worsham (2009) warns persons making decisions based on data from Affinity 2.5, recommending that clinicians administer this assessment at least twice to the same individual in order to check the temporal stability of the person's response patterns (regardless of what the patterns are—typical or deviant) before drawing conclusions or planning treatments. Ipsative logic supports this suggestion.

Alternatively, could sexual interest itself or perception of sexual attractiveness or both be generally unstable across time? Is it possible that researchers may have operationalized sexual interest in such a way that it is prone to appear to vary from T1 to T2 when it is being tested? If sexual interest is unstable, what precautions are appropriate for making decisions based on assessment of a construct that is unstable among a noticeable percentage of participants? Affinity

2.5 purports to measure sexual interest at a specific point in time. It seems wise to be cautious when interpreting results from any data collected using this tool so as to avoid extrapolating beyond propriety any predictions of past or future sexual interest.

Could it be that the context in which the person completed the Affinity (with whom he or she had interacted that day, how he or she was feeling, about what he or she was thinking prior to completing the assessment) may affect the person's speed for creating and enacting schemas to facilitate categorization of the persons depicted in the assessment images? Indeed, Fiske and Neuberg (1990) alluded to such a possibility with their assertion that people only sometimes use cognitive shortcuts when judging others, and at other times attend diligently to the data and process them carefully. It is a viable possibility that a person's sexual interest or perception of attractiveness may actually change based on his or her context, or may appear to change based on his or her employment of schemas. What are the implications of this possibility, and how might one research them? It seems appropriate to concede that while the Affinity does not provide data that reflects Truth, it may yield viewing time scores that reflect a sort of truth defined by context.

Is it possible that the assessment tool itself is unstable and therefore likely to report data reflecting this instability? If the tool is unstable, how is it useful as a measure of sexual interest? Perhaps answers to these questions require a reconceptualization of the concept of reliability itself. Reliability is often misconstrued as a principle of psychometric and technical adequacy simply applied to a test. Reliability might more appropriately be understood as an artifact of the testing *context*, not only the test, and reflect the influences on the test and the people who take it. Perhaps temporal stability can be better conceptualized as a principle determined on a case by case basis. In light of these questions, it seems important to note that for all their instability, the Affinity instruments themselves yield fairly consistent percentages of stable, nonsignificant chi-

square values (see Table 2). This trend is notable since efforts to determine the assessment tool's validity are taking into consideration the tool's ability to repeatedly report similar patterns of scores for similar samples of participants.

Study Limitations

The limitations of this study are similar to those of other tests of the reliability of scores for assessment tools. This study is limited by the fact that each participant completed the assessment only twice and typically participated in the second administration within several weeks (but no less than 14 days) of completing the test the first time. It is unknown how patterns of responding may or may not have differed had participants taken the test more than twice, or if a longer or shorter amount of time had elapsed between test administrations.

Another limitation of this study is based in its foundational task of using viewing time as a measure of sexual interest. While there is important literature and theory to support this premise, Quinsey and colleagues' (1996) caution is worth repeating: Because visual appraisal can also serve functions unrelated to sexuality, viewing time is prone to influence by procedural and instructional variables which may affect the data collected from the measure. While efforts were made to standardize for participants as many procedural and instructional components as feasible, it is possible that variations in test administration may have affected the results.

The generalizability of the results of this study is also limited by the demographics of the participants. For this study, the sample consisted of participants largely from one ethnic group, all students at a large, private, religious university, all exclusively heterosexual with no history of pedophilia, and all recruited voluntarily. As described in the review of literature, there is evidence to suggest that volunteers for participation in studies involving issues of sexuality may differ from their nonvolunteering peers on specific factors (Strassberg & Lowe, 1995; Suschinsky et al., 2009; Wolchik et al., 1985). Thus, it is probable that biases exist in this

sample. Even so, the results of this study provide valuable information about the temporal stability of Affinity 2.5 for a group of people with specific characteristics. Ethical practice requires that researchers, clinicians, and others consider the composition of this particular sample when interpreting the results of this study.

Future Research

This study adds to the existing body of research for Affinity 2.5 an important reliability component in the quest to examine this tool's validity, or whether it actually measures what it purports to measure. Analyses of temporal stability data support the reliability of scores on Affinity 2.5. Normative response patterns for exclusively heterosexual, nonpedophilic males and females have been established previously (Boardman, 2009; Worsham 2009) and allow for comparison of responses on Affinity 2.5 for other people who fit this description. However, there are other pedophilic and nonpedophilic groups whose normative response patterns will likely differ from the normative patterns established by these earlier studies. Normative patterns for Affinity 2.5 might be investigated for persons with different sexual orientations, sexual preferences, ethnicities, genders, age groups, and other pertinent identifying characteristics in order to establish a store of reference groups against which future participants' proportion means might be compared.

In addition to expanding the variety of normative samples for Affinity 2.5, this measure shows promise as a screening tool for sexual deviancy in clinical, correctional, and judicial settings, with the important stipulation that any screening be conducted twice for each individual being assessed in order to inspect the contextual stability of the person's pattern of responses. If assessment tools are being used to inform clinical and legal decisions that affect a person's life, it is important that the information provided by the instrument be reliable, valid, and interpreted correctly.

Behaviors related to pedophilia qualify as sexual deviance, as defined by Hanson and Bussière (1998), because they are illegal and deviate from socially and legally acceptable norms. As part of her argument for the necessity of creating valid and reliable measures of sexual interest, Gress (2005) stressed the importance of developing an assessment tool for discovering sexually deviant preferences, especially for pedophilia. Results from such assessment tools might be used with sexual offenders to determine treatment needs, to identify high-risk situations, and to predict the possible rate of recidivism (Fischer, 2000; Gress, 2005; Marshall, 1996). In addition, such an assessment tool may be used with persons convicted of sexual offenses in their youth who do not actually manifest sexual preferences for children. Affinity 2.5 may be instrumental in separating persons with true deviant sexual interests in children from persons convicted of crimes of convenience involving minors. Such distinctions seem important to consider when addressing issues related to treatment planning, judicial rulings, and public records.

Research using Affinity 2.5 is ongoing and promising, and its application to various settings and clinical situations continues to advance. While its intended purpose and design is in clinical work with sex offenders, the relatively high reliability of its scores across time examined by the current study supports its validity as a measure with potentially wider application.

References

Abel, G. G. (1996). *A new objective test for youthful offenders: The Abel Assessment for Interest in Paraphilias*. Atlanta, GA: Abel Screening Inc.

Abel, G.G. (1997). Abel Assessment for Sexual Interest: Judges' product information. Atlanta, GA: Abel Screening Inc.

Abel, G.G., Coffey, L., Osborn, C.A. (2008). Sexual arousal patterns: Normal and deviant. *The Psychiatric Clinics of North America, 331*, 643-655.

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.

Abel, G. G., Lawry, S. S., Karlstrom, E. M., Osborn, C. A., & Gillespie, C. F. (1994). Screening tests for pedophilia. *Criminal Justice and Behavior*, *21*, 115–131.

Abel, G.G., Jordan, A., Rouleau, J.L., Emerick, R., Barboza-Whitehead, S., & Osborn, C. (2004). Use of visual reaction time to assess male adolescents who molest children. *Sexual Abuse: A Journal of Research and Treatment, 16*(3), 255-265.

Abel, G.G., Mittelman, M.S., & Becker, J.V. (1985). Sexual offenders: results of assessment and recommendations for treatment. In Ben-Aron, H., Hucker, S.I., & Webster, C.D. (Eds.). *Clinical Criminology* (pp. 191-205). Toronto, Canada : M.M. Graphics.

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

Allport, G.W. (1954). *The nature of prejudice.* Cambridge: Addison-Wesley.

Amoroso, D.M., Brown, M., Pruesse, M., Ware, E.E., & Pilkey D.W. (1970). An investigation of behavioral, psychological and physiological reactions to pornographic stimuli.

Technical Reports of the Commission on Obscenity and Pornography, 8. Washington, D.C.: U.S. Government Printing Office.

Anastasi, A. (1988). *Psychological testing.* New York: Macmillan.

Arnheim, R. (1954). *Art and visual perception: A psychology of the creative eye.* Berkeley: University of California Press.

Arnheim, R. (1966). *Toward a psychology of art.* Berkeley: University of California Press.

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

Barbaree, H.E., Mewhort, K.J.K. (1994). The effects of the *z*-score transformation on measures of relative erectile response strength: A re-appraisal. *Behaviour Research and Therapy, 32,* 547-558.

Barrett, P. (2003). Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology, 18*(5), 421-439.

Bartlett, F.A. (1932). *Remembering: A study in experimental and social psychology.* New York: Cambridge University Press.

Berscheid, E., & Walster, E. (1974). Physical attractiveness. In L. Berkositz (Ed.). *Advances in experimental social psychology* (pp. 157-215). New York: Academic Press.

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

Boardman, R.D. (2009). *Creating an expected profile for Affinity 2.5 from a sample of non-pedophilic, exclusively heterosexual, college-age males.* Unpublished doctoral dissertation, Brigham Young University, Provo, Utah.

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., Preuss, M., & Pilkey, D.W. (1973). Factors affecting viewing time of pornography. *Journal of Social Psychology, 90,* 125-135.

Bullock, D.H. (1959). Note on "looking at pictures" behavior. *Perceptual Motor Skills, 9,* 333.

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

Cavior, N., & Lombardi, D.A. (1973). Developmental aspects of judgment of physical attractiveness in children. *Developmental Psychology, 8*(1), 67-71.

Child, I.L. (1962). Personal preferences as an expression of aesthetic sensitivity. *Journal of Personality, 30,* 496-512.

Child, I.L., & Cooperman, M. (1968). Relation of esthetic sensitivity to psychoanalytic character traits. *Proceedings of the 76th Annual Convention of the American Psychological Association, 3,* 439-440.

Child, I.L., Hansen, J.A., & Hornbeck, F. (1968). Age and sex differences in children's color preferences. *Child Development, 39,* 237.

Child, I.L., & Kasti, A.J. (1968). Comparison of color preferences in Vietnam and the U.S. *Proceedings of the 76th Annual Convention of the American Psychological Association, 3,* 437-438.

Chivers, M.L., Rieger, G., Latty, E., & Bailey, J.M. (2004). A sex difference in the specificity of sexual arousal. *Psychological Science, 15,* 736-744.

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

Cloyd, L.L. (2007). *Relationship between key variables in penile plethysmography and viewing time measures of sexual arousal in sex offending adult males.* Unpublished doctoral dissertation, University of Miami, Coral Gables, Florida.

Crosby, D. (2007). *Non-pedophilic heterosexual male response to the Affinity 2.0.* Unpublished doctoral dissertation, Brigham Young University, Provo, Utah.

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

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

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

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

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,* 209-241.

Ewing, C.P. (2006, January). Testing tool in question. *Monitor on Psychology, 37*(1), 61.

Finkelhor, D., Ormrod, R., Turner, H., & Hamby, S. L. (2005). The victimization of children and youth: A comprehensive, national survey. *Child maltreatment: Journal of the American Professional Society on the Abuse of Children, 10*(1).

Fischer, L. (2000). The Abel Screen: A non-intrusive alternative? In D.R. Laws, S.M. Hudson, & T. Ward (Eds.), *Remaking relapse prevention with sexual offenders: A sourcebook* (pp. 303-318). Thousand Oaks: Sage.

Fischer, L. (2004, October). Psychometric issues in viewing-time research, clinical, and forensic applications. In L. Fischer (Chair), *Accessible, standardized, viewing-time measurement*. Symposium conducted at the 23rd Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Albuquerque, NM.

Fischer, L. (2006a). *Sex offenders' response to Affinity 2.0*. Manuscript in progress.

Fischer, L. (2006b, September). Norm-referenced clinical decision making with Affinity viewing time results. In L. Fischer (Chair), *Current research and clinical application of Affinity*. Symposium to be conducted at the 25th annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL.

Fischer, L., Byrne, P., & Glasgow, D. (2007, October) Real world decision-making using both ipsative and norm-referenced Monarch PPG and Affinity results. Paper presented at the 26th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Fischer, L. & Meade, A.W. (2010). Assessment of sexual interests: Reference-group solutions to problems of ipsative data. Manuscript in progress.

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 adequacy of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11*(3), 195-205.

Fiske, S.T., & Neuberg, S.L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and

interpretation. In M.P. Zanna (Ed.). *Advances in experimental social psychology,* (vol. 23, pp. 1-74). New York: Academic Press.

Fiske, S.T., & Taylor, S.E. (1991). *Social cognition.* 2nd ed. Reading, MA: Addison-Wesley.

Freund, K., Langevin, R., & Zajac, Y. (1974). Heterosexual aversion in homosexual males. *British Journal of Psychiatry, 122,* 163-169.

Freund, K., Langevin, R., & Zajac, Y. (1974). Heterosexual aversion in homosexual males: A second experiment. *British Journal of Psychiatry, 125,* 177-180.

Gaither, G.A., & Plaud, J.J. (1997). The effects of secondary stimulus characteristics on men's sexual arousal. *The Journal of Sex Research, 34*(3), 231-236.

Glasgow, D. (2006, September). Viewing-time as a measure of social cognition. 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.

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

Glasgow, D. & Croxen, J. (2003). Viewing time as a measure of paedophile sexual interest: Empirical evidence, the potential of social cognitions as a theoretical model, and some possible implications for assessment and intervention. Unpublished manuscript.

Glasgow, D.V., Osborne, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: An application of single case methodology. *British Journal of Learning Disabilities, 31,* 96-102.

Glasgow, D., & Fischer, L. (2006a, June). Viewing-time and sex offender assessment: Psychometric issues & possible psychological mechanisms. Paper presented at the 16th Conference of the European Association of Psychology & Law (EAPL), Liverpool, UK.

Glasgow, D., & Fischer, L. (2006b, September). Assessment of paedophile sexual interest: Theoretical and practical issues and the potential of viewing time (VT) assessments. Paper presented at the 16th Annual Conference and Practice Workshops of the National Organisation for the Treatment of Abusers (NOTA), York, UK.

Gress, C.L.Z. (2001). An evaluation of a sexual interest assessment tool. Unpublished manuscript.

Gress, C.L.Z. (2005). Viewing time measures and sexual interest: Another piece of the puzzle. *Journal of Sexual Aggression, 11*(2), 117-125.

Hamilton, D.L., & Sherman, S.J. (1994). Stereotypes. In R.S. Wyer, Jr., & T.K. Srull (Eds.). *Handbook of social cognition* (vol. 2, pp. 1-68). Hillsdale, NJ: Erlbaum.

Hanson, R.K., & Bussière, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*(2), 348-362.

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

Harris, G.T., Rice, M.E., Quinsey, V.L., & Chaplin, T.C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behavior Research and Therapy, 34*(4), 389-394.

Hassebrauck, M. (1998). The visual process method: A new method to study physical attractiveness. *Evolution and Human Behavior, 19,* 111-123.

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

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

Howard, J.A., & Renfrow, D.G. (2003). Social cognition. In J. Delamater (Ed.). *Handbook of social psychology* (pp. 259-281). New York: Kluwer Academic/Plenum Publishers.

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

Howes, R.J. (2003). Circumferential change scores in phallometric assessment: Normative data. *Sexual Abuse: A Journal of Research and Treatment, 15*(4), 365-375.

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

Israel, E., & Strassberg, D.S. (2009). Viewing time as an objective measure of sexual interest in hetersexual men and women. *Archives of Sexual Behavior, 38,* 551-558.

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

Kirschner, N.M. (1976). Effect of need for approval and situational variables on the viewing of erotic material. *Journal of Consulting and Clinical Psychology, 44,* 869.

Krueger, R.B., Bradford, J.M.W., & Glancy, G.D. (1998). Report from the Committee on Sex Offenders: The Abel Assessment for Sexual Interest—A brief description. *Journal of American Academy of Psychiatry and the Law, 26*(2) 277-280.

Lang, A.R., Searles, J., Lauerman, R., & Adesso, V. (1980). Expectancy, alcohol, and sex guilt as determinants of interest in and reaction to sexual stimuli. *Journal of Abnormal Psychology, 89*(5), 644-653.

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

Langlois, J.H., Roggman, L.A., Casey, R.J., Ritter, J.M., Rieser-Danner, L.A., & Jenkins, V.Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology, 23,* 363-369.

Langlois, J.H., Roggman, L.A., & Rieser-Danner, L.A. (1990). Infants' differential social responses to attractive and unattractive faces. *Developmental Psychology, 26*(1), 153-159.

Laws, D.R. (1989). *Relapse prevention with sexual offenders.* New York: Guilford.

Laws, D.R. & Gress, C.L.Z. (2003). Inter-rater reliability study of Not Real People (NRP) Set A. Unpublished raw data.

Laws, D.R., & Gress, C.L.Z. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9,* 183-196.

Laws, D.R., Hanson, R.K., Osborn, C.A., & Greenbum, P.E. (2000). Classification of child molesters by plethysmographic assessment of sexual arousal and a self-report measure of sexual preference. *Journal of Interpersonal Violence, 15,* 1297-1312.

Leckart, B.T., Kelling, K.R., & Bakan, P. (1966). Sex differences in the duration of visual attention. *Perception & Psychophysics, 1,* 374-376.

Letourneau, E.J. (2002). A comparison of objective measures of sexual arousal and interest: Visual reaction time and penile plethysmography. *Sexual Abuse: Journal of Research and Treatment, 14*(3), 207-223.

Linville, P.W., Salovey, P., & Fischer, G.W. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality & Social Psychology, 57*(2), 165-188.

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

Macrae, C.N., & Bodenhausen, G.V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology, 92,* 239-255.

Madsen, J.B. (2008). *Males' ipsative score distortion on Affinity 2.0.* Unpublished specialist thesis, Brigham Young University, Provo, Utah.

Marshall, W.L. (1996). Assessment, treatment, and theorizing about sex offenders: Developments during the past twenty years and future directions. *Criminal Justice and Behavior, 23,* 162-199.

Marshall, W.L., & Fernandez, Y.M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review, 20*(7), 807-822.

Martin, B. (1964). Expression and inhibition of sex motive arousal in college males. *Journal of Abnormal and Social Psychology, 68*(3), 307-312.

Meston, C.M., Heiman, J.R., Trapnell, P.D., & Paulhus, D.L. (1998). Socially Desirable Responding and Sexuality Self-Reports. *The Journal of Sex Research, 35*(2), 148-157.

Mosher, D.L. (1966). The development and multitrait-multimethod matrix analysis of three measures of three aspects of guilt. *Journal of Consulting Psychology, 30,* 25-29.

Mueser, K.T., Grau, B.W., Sussman, S., & Rosen, A.J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology, 46*(2), 469-478.

Pacific Psychological Assessment Corporation (PPAC). (2004). The NRP (Not Real People) stimulus set for assessment of sexual interest. Victoria, BC, Canada: Author.

Polan, M.L., Desmond, J.E., Banner, L.L., Pryor, M.R., McCallum, S.W., Atlas, S.W., et al. (2003). Female sexual arousal: A behavioral analysis. *Fertility and Sterility, 80*(6), 1480-1487.

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

Quinsey, V.L., & Chaplin, T.C. (1988). Preventing faking in phallometric assessment of sexual preference. *Annals of the New York Academy of Sciences, 528,* 49-58.

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

Quinsey, V.L., Rice, M.E., Harris, G.T., & 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* (pp. 143-163). New York: Guilford.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.). *Cognition and categorization (*pp. 27-48). Hillsdale, NJ: Erlbaum.

Rosenzweig, S. (1942). The photoscope as an objective 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.

Rupp, H.A., Librach, G.R., Feipel, N.C., Ketterson, E.D., Sengelaub, D.R., & Heiman, J.R. (2009). Partner status influences women's interest in the opposite sex. *Human Nature, 20,* 93-104.

Rupp, H.A., & Wallen, K. (2009). Sex-specific content preferences for visual sexual stimuli. *Archives of Sexual Behavior, 38,* 417-426.

Samuels, C.A., & Ewy, R. (1985). Aesthetic perception of faces during infancy. *British Journal of Developmental Psychology, 3,* 221-228.

Santtila, P., Mokros, A., Viljanen, K., Koivisto, M., Sandnabba, N.K., Zappalà, A., et al. (2009). Assessment of sexual interest using a choice reaction time task and priming: A feasibility study. *Legal and Criminological Psychology, 14,* 65-82.

Saunders, R.M., & Naus, P.J. (1993). The impact of social content and audience factors on responses to sexually explicit videos. *Journal of Sex Education and Therapy, 19*(2), 117-130.

Schiffer, B., Krueger, T., Paul, T., de Greiff, A., Forsting, M., Leygraf, N., et al. (2008). Brain response to visual sexual stimuli in homosexual pedophiles. *Journal of Psychiatry and Neuroscience, 33*(1), 23-33.

Shapiro, B.A., Eppler, M., Haith, M.M., & Reis, H. (1987, April). *An event analysis of facial attractiveness and expressiveness.* Paper presented at the meeting of the Society for Research in Child Development, Baltimore, MD.

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 sexual offenders: Reliability and validity of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11*(3), 207-216.

Snyder, H. (2000, July). *Sexual assault of young children as reported to law enforcement: Victim, incident, and offender characteristics*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Snyder, H., & Sickmund, M. (2006). *Juvenile offenders and victims: 2006 national report*. Washington, D.C.: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

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

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

Suschinsky, K.D., Lalumière, M.L., & Chivers, M.L. (2009). Sex differences in patterns of genital sexual arousal: Measurement artifacts or true phenomena? *Archives of Sexual Behavior, 38,* 559-573.

Tajfel, H. (1969). Social and cultural factors in perception. In G. Lindzey & E. Aronson (Eds.). *Handbook of social psychology* (vol. 3, pp. 315-394). Reading, MA: Addison-Wesley.

U.S. Census Bureau. (2011). *Statistical abstract of the United States: 2011.* Retrieved from http://www.census.gov/compendia/statab/2011/tables/11s0310.pdf

United States Department of Justice (2006) *Juvenile Offenders and Victims: 2006 National Report*. Retrieved January 18, 2010 from http://ojjdp.ncjrs.gov/ojstatbb/nr2006/downloads/NR2006.pdf.

Valentine, C.W. (1962). *The experimental psychology of beauty.* New York: Dover.

Vanstone, K., & Laws, D.R. (1998). Qualitative and Attractiveness Rating Scales. Unpublished manuscript, Adult Forensic Psychiatric Community Services Victoria, BC, Canada.

Ward, T., Louden, K., Hudson, S.M., & Marshall, W.L. (1995). A descriptive model of the offense chain for child molesters. *Journal of Interpersonal Violence, 10,* 342-372.

Ware, E.E., Brown, M., Amoroso, D.M., Pilkey, D.W., & Preuss, M. (1972). The semantic meaning of pornographic stimuli for college males. *Canadian Journal of Behavioral Science, 4,* 204-209.

Warner, R. (2008). *Applied Statistics: From Bivariate through Multivariate Techniques.* Los Angeles: Sage Publications.

Whitaker, D.J., Le, B., Hanson, R.K., Baker, C.K., McMahon, P.M., Ryan, G., et al. (2008). Risk factors for the perpetration of child sexual abuse: A review and meta-analysis. *Child Abuse & Neglect, 32,* 529-548.

Wincze, J.P., Hoon, E.F., & Hoon, P.W. (1972). Multiple measure analysis of women experiencing low sexual arousal. *Behavioral Research and Therapy, 16,* 43-49.

Wolchik, S.A., Braver, S.L., & Jensen, K. (1985). Volunteer bias in erotica research: Effects of intrusiveness of measure and sexual background. *Archives of Sexual Behavior, 14,* 93-107.

Worling, J.R. (2006). Assessing sexual arousal with adolescent males who have offended sexually: Self-report and unobtrusively measured viewing time. *Sexual Abuse: A Journal of Research and Treatment, 18,* 383-400.

Worsham, M. (2009). *Creating an expected profile for Affinity 2.5 from a sample of non-pedophilic, exclusively heterosexual, college-age females.* Unpublished doctoral dissertation, Brigham Young University, Provo, Utah.

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.

Zebrowitz, L.A. (1996). Physical appearance as a basis of stereotyping. In C.N. Macrae, C. Stangor, & M. Hewstone (Eds.). *Stereotypes and stereotyping* (pp. 79-120). New York: Guilford.

Zerubavel, E. (1996). Lumping and splitting: Notes on social classification. *Sociological Forum, 11*(3), 421-433.

Appendix A

Consent to be a Research Subject

Introduction
This research study is being conducted by Kristina Hansen, Ph.D. student, and Lane Fischer, Ph.D., at Brigham Young University to examine the temporal stability of responses to Affinity 2.5 by adult males and females. You were selected to participate because you are over age 18 and have no history of pedophilia.

Procedures
You will be asked to complete the Affinity 2.5 in a private room in the CPSE research lab (350 MCKB). Affinity 2.5 is a computer-administered measure of sexual interest. You will be asked to rank order some line drawings of types of people according to their sexual attractiveness and unattractiveness to you. You will then be asked to rate a series of images of clothed models in everyday activities according to how sexually attractive or unattractive they are to you. No pornographic images are used in Affinity 2.5. Following completion of the Affinity 2.5, you will be asked to fill out a brief questionnaire regarding some simple demographics, personal attitudes and sexual preference. You will return and complete the Affinity 2.5 again in approximately 14 days. The procedure will take approximately 20 minutes to complete each time.

Risks/Discomforts
There are minimal risks for participation in this study. However, you may feel some discomfort about disclosing sexual interests or rating images of people. The possibility of a breach of confidentiality of potentially sensitive information regarding sexual preferences will be mediated by use of subject ID numbers, keeping this signed consent form unconnected to responses to Affinity 2.5 or the questionnaire, and limiting researcher access to consent forms and data connected to participants.

Benefits
There are no direct benefits to you. However, it is hoped that through your participation researchers will learn more about how people respond to such rating tasks and help us better understand human sexuality.

Confidentiality
All information provided will remain confidential. Your responses will be assigned a subject number that will be disconnected from your name. Your responses will be downloaded from Affinity 2.5 to Excel and another statistical program and then erased from the Affinity program files. The questionnaire will also be coded only by a subject number, transcribed into Excel and SPSS and separated from your name. After the research is completed, the questionnaires will be destroyed. Although the questionnaire will ask about your sexual preference, no information will be available to the university or the Honor Code Office.

Compensation
Participants may receive extra credit or clinical hours in their classes that offer such compensation. An alternative method of compensation may be provided at the discretion of your instructor, and often consists of reviewing a journal article or some other activity which requires a time commitment similar to participating in the current study.

Participation
Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without jeopardy to your class status, grade or standing with the university.

Questions about the Research
If you have questions regarding this study, you may contact Kristina Hansen at (801) 703-5934, kristina_hansen@byu.edu or Lane Fischer at (801) 422-8293, lane_fischer@byu.edu .

Questions about your Rights as Research Participants
If you have questions you do not feel comfortable asking the researcher, you may contact BYU IRB Administrator, A-285 ASB, Brigham Young University, Provo, UT 84602, 801-422-1461, irb@byu.edu,
****

I have read, understood, and received a copy of the above consent and desire of my own free will to participate in this study.

Signature:_____ Date:_____

Appendix B

Demographics, Attitudes, and Sexual Interest Questionnaire

Subject #

Demographics

1. Age: ____

2. Ethnicity: _____

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

   ___Freshman ___Sophomore
   ___Junior ___Senior
   ___Graduate Student

4. Marital Status (mark the one that applies):

   ___Single ___Married
   ___Divorced ___Widowed

Personal Attitudes

5. 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

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

_____ Exclusively heterosexual with no homosexual interest

_____ Predominantly heterosexual with incidentally homosexual interest

_____ Predominantly heterosexual with more than incidentally homosexual interest

_____ Equally heterosexual and homosexual interest

_____ Predominantly homosexual with more than incidentally heterosexual interest

_____ Predominantly homosexual with only incidentally heterosexual interest

_____ Exclusively homosexual with no heterosexual interest

Appendix C

Chi-Square (χ^2) Calculations of Affinity 2.0 Data Using Constant $n = 115$

Participant No.	χ^2	Participant No.	χ^2
3001	2.678784483	3045	9.543295455
3002	5.047222222	3046	4.174848485
3003	1.39482018	3047	0.9321533
3004	5.568519814	3048	8.38922646
3005	7.866659918	3050	3.612792398
3006	5.006420455	3052	5.70852591
3007	4.595588715	3053	44.80632901 *
3008	14.69732112 *	3054	31.41535594 *
3009	3.539859307	3055	70.51330728 *
3011	19.50104618 *	3056	6.798887427
3012	23.38456328 *	3057	5.41631746
3013	23.35478355 *	3058	11.09120047
3014	21.26033751 *	3059	3.051515873
3015	8.858173701	3060	7.270769537
3016	2.022318422	3061	23.61242063 *
3017	3.000099206	3062	28.60670635 *
3018	9.29370124	3063	7.57008658
3019	1.819090909	3065	17.52118869 *
3020	21.55680546 *	3066	3.258573517
3021	1.729035714	3073	7.872625361
3022	2.669109649	3075	14.66684078 *
3023	10.02596278	3076	11.4554533
3024	4.394794974	3077	10.52757353
3025	2.881244779	3078	10.47263986
3026	5.221132756	3081	13.45241402
3027	4.504166667	3082	7.253550543
3028	4.651909722	3083	18.333653385 *
3029	9.358809524	3084	8.594652778
3030	6.40128113	3090	7.588455882
3032	3.216779932	3091	1.878082789
3033	15.69339286 *	3092	1.149663036
3034	5.832142857	3093	0.660959596
3036	6.111693548	3096	5.735504202
3037	4.376388889	3097	2.027156863
3038	4.165265152	3102	1.249027778
3041	5.360825397	3104	5.553061299
3042	7.389753968	3106	7.538450092
3043	8.577859934	3109	13.07651099
3044	3.598175204		

Note. χ^2 values were calculated from viewing time proportion data reported in Crosby (2007).

* p >.05 significant critical value (14.067).

Appendix D

Affinity 2.5 Chi-Square Results for a Sample of Exclusively Heterosexual, Nonpedophilic Males

Participant No.	χ^2	Participant No.	χ^2
7001	1.976022822	7047	1.462055606
7002	0.604306998	7048	6.523432374
7004	3.056709659	7049	6.485688042
7006	3.715765895	7050	13.49467899
7011	4.631914027	7051	3.461097645
7014	2.349151036	7052	13.92636956
7016	48.63815591 *	7053	7.982800501
7017	4.731518271	7054	14.77802992 *
7020	15.03764043 *	7055	11.82889267
7021	9.656779862	7058	3.034575175
7022	6.247729618	7061	14.47840509 *
7024	4.136431922	7062	2.860787117
7027	8.589368183	7063	10.82268285
7028	7.357777629	7065	11.4902736
7029	6.030185018	7066	6.647431234
7030	9.255479019	7067	10.84824274
7031	6.576346692	7069	2.286482375
7032	1.557960031	7070	7.771364334
7033	9.69167198	7071	9.593015505
7034	3.228175692	7073	14.21988447 *
7035	2.616963835	7074	2.060620563
7036	12.47823625	7075	6.083768834
7037	3.168257964	7077	12.99575474
7038	3.497486867	7078	21.66133421 *
7039	4.624813056	7079	3.273266234
7040	7.067693199	7080	14.53046035 *
7041	7.549127747	7082	3.294147239
7042	12.66169055	7083	10.15968226
7043	42.66006014 *	7084	4.264120907
7044	46.33971004 *	7086	6.323736031
7045	4.64182439	7087	3.276094741
7046	11.1781448		

* p >.05 significant critical value (14.067).

Appendix E

Affinity 2.5 Chi-Square Results for a Sample of Exclusively Heterosexual, Nonpedophilic Females

Participant No.	χ^2	Participant No.	χ^2
8002	37.72168655 *	8054	4.832372677
8003	23.7342587 *	8055	6.026046197
8004	3.839645161	8056	5.167468302
8005	3.467000632	8057	2.779424768
8006	1.499958038	8058	11.65372701
8007	1.156425587	8060	3.630140713
8009	7.75562502	8061	10.37337132
8010	4.973895539	8064	6.412648962
8011	3.18275976	8065	3.74380568
8012	6.881125239	8066	9.120507112
8013	4.515541683	8067	5.084838189
8014	2.314670186	8069	3.622194768
8015	9.145360489	8070	1.88550924
8018	19.38392555 *	8072	7.679165033
8019	2.018449226	8073	7.806909824
8021	3.456330343	8074	4.364839461
8023	1.27793547	8075	2.144430324
8025	7.760766014	8076	6.672756487
8028	3.643277694	8077	4.004132144
8029	2.035696805	8078	4.026315957
8030	7.064625472	8080	5.083995839
8031	6.642249177	8081	1.719850291
8032	5.630281946	8082	8.681982564
8033	12.15222093	8083	3.565515038
8034	62.64257274 *	8085	1.449489945
8035	5.163801052	8086	15.23532201 *
8036	12.01002458	8087	11.44071593
8037	5.118691026	8088	7.96082544
8038	6.300568883	8091	16.4016491 *
8039	2.154191077	8092	3.208057005
8040	0.617103214	8093	4.321898262
8041	1.871628293	8094	12.90894315
8042	53.74831395 *	8095	13.3092484
8043	11.49554572	8096	3.279117517
8044	23.79309801 *	8097	5.473470957
8046	1.016547385	8098	4.248548487
8047	14.17143972 *	8100	11.9785916
8048	4.964716953	8102	4.139107396
8049	4.171139065	8103	4.737393822
8050	0.875575527	8105	6.773822922
8051	10.08880392	8106	2.564759444
8053	14.19032999 *	8107	1.403504995

* $p > .05$ significant critical value (14.067).