2014-08-28

Screening and Diagnostic Validity of Affinity 2.5

Heather Lynn Stephenson
Brigham Young University - Provo

Screening and Diagnostic Validity of Affinity 2.5

Heather Stephenson

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
Aaron P. Jackson
Ellie L. Young
Marleen S. Williams
P. Scott Richards

Department of Counseling Psychology and Special Education

Brigham Young University

August 2014

ABSTRACT

Screening and Diagnostic Validity of Affinity 2.5

Heather Stephenson
Department of Counseling Psychology and Special Education, BYU
Doctor of Philosophy

Affinity 2.5 is a computer-based instrument designed to assess sexual interest using viewing-time measures. Viewing-time measures of sexual interest have been developed to identify individuals with deviant sexual interest. The purpose of this study is to examine the validity of Affinity 2.5 in screening and diagnosing individuals with sexually deviant interests. This study used viewing time profiles of known sexual offenders compared to norm-referenced profiles of an exclusively heterosexual, non-pedophilic college population. Participants were 155 males and 3 females who had sexually offended against children and 63 male and 84 female non-offender college students. Results show that 43.7% of offenders were correctly identified as having significantly deviant sexual interest, compared to the reference group. Further 12.0% of offenders showed statistical significant interest in at least one category of individuals from a protected population and offended against that same category. The results of this study do not provide support for the utility of the Affinity 2.5 as a screening or diagnostic tool.

Keywords: Affinity 2.5, sexual interest, viewing time, reliability, validity, chi-square

ACKNOWLEDGMENTS

This project could not have been possible without the help and support of those around me. Special thanks to Lane Fischer for his patience, guidance, tenacity, and ability to rein in my desire for absolute truths. I will hold with fondness the memory of the hours and hours we hit our heads against the wall working through statistical roadblocks and problems. Thank you for buffering the stresses and dead ends that presented themselves during this project.

I would like to thank my committee members for encouraging me to pursue this study and working with me through all the iterations. Thanks also to the Affinity team for support, assistance, and the tireless debates on methods and ethics. Special thanks to Kristina Hansen, Sierra Baird, and Joy Cox for your efforts in seeing this through with me. I want to express my gratitude to Mervyn Davies for his time, office, resources, and friendship.

I am grateful to the Brigham Young University Graduate School for awarding me the Graduate Research Fellowship that allowed this project to see a speedy resolution and opened the door for other research endeavors.

Sincerest thanks go to my family who has never wavered in support of me while attaining my degree. Special thanks to my husband Brandon for sacrificing time for my long statistics and writing sessions. He is my best friend and my happiness. Without him, I am incomplete.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF DISSERTATION STRUCTURE

This dissertation, *Screening and Diagnostic Validity of Affinity 2.5*, is written in a hybrid format. The hybrid format is designed to combine traditional dissertation requirements with a journal-ready format. The initial pages are requirements for submission to the university. The dissertation study is written in a journal-ready format which conforms to length and style requirements of most psychological journals. This is followed by a literature review located in Appendix A. This dissertation format includes two reference lists. The first includes those references of the journal-ready article. The second list comprises all references and citations used in the review of literature found in the Appendix.

Introduction

The National Center for Victims of Crime (2010) report states that 6.1% of children under the age of 17 are victims of sexual crimes. More than half of child victims of sexual abuse were abused by individuals known to them (53-66%), and over 2,900 incidents for viewing juvenile pornography were reported in the year 2000. Over the course of their lifetime, 28% of 14-17 year olds will experience some form of sexual abuse (The National Center for Victims of Crime, 2010). A 1988 national sample of adult men in the United States revealed that 17% of these men reported having sexually molested a child (Finkelhor & Lewis, 1988). These staggering statistics reveal that many sexual offenders of children remain unidentified and this has fueled much research into sexual development, deviance, and trauma recovery (Abel, Lawry, Karlstrom, & Osborn, 1994; Abel et al., 2004; Brown, 1979; Cunningham, 1986; Gaither & Plaud, 1997; Gannon, Rose, & Ward, 2008; Glasgow, Osborne, & Croxen, 2003; Hanson & Bussiere, 1998; Kinsey, Pomeroy, & Martin, 1998; Kinsey, Pomeroy, & Martin, 2003).

One area of research regarding sexual interest utilizes viewing-time measures of sexual interest. Viewing-time (VT) instruments are used based on the components of sexual arousal that were proposed by Singer (Singer, 1984). Specifically, they address the aesthetic response of Singer's model with the assumption that individuals, when asked to rate the relative sexual attractiveness of an individual from an image, will spend more time viewing the image if the object portrayed is comparable to the individual's ideal sexual partner (Rosenzweig, 1942). When asked to rate the sexual attractiveness of images that an individual considers highly unattractive, the respondent is less likely to spend time cognitively processing the stimulus, and will instinctively respond (Macrae & Bodenhausen, 2000). When asked to rate images that are closer to the respondent's ideal target of sexual attraction, the respondent must cognitively

process the image in order to classify and rate the image (Macrae & Bodenhausen, 2000; Macrae & Bodenhausen, 2001). VT measures have been successful in discriminating between heterosexual males and gay males (Rosenzweig, 1942), heterosexual men and women (Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996), individuals with and without pedophilia (Harris, Rice, Quinsey, & Chaplin, 1996a), and individuals who experience high guilt around sexual topics versus those who do not (Love, Sloan, & Schmidt, 1976). Using Affinity 2.5, Mokros et al. (2013) was able to discriminate between sexual offenders, non-sexual offenders, and non-offending controls.

A major concern in utilizing viewing time measures, however, is that the data are ipsative, and all data sum to a constant (Cattell, 1944; Hicks, 1970; Radcliffe, 1963). When using ipsative data, there are a number of limitations in statistical analysis and interpretation. The first, and perhaps most significant limitation of ipsative data is one of non-independence. Statistical analyses require observations to be statistically independent or the covariance matrix is considered singular. When the covariance matrix is singular, this means that the data are co-dependent or lacking in variance, which requires the data to be modified or one variable omitted to produce results. However, this modification invariably changes the results. Therefore, standard forms of statistical analyses (e.g., structural equation modeling) are inappropriate when using ipsative data.

A second limitation of ipsative data is that it is essentially an ordinal level of measurement and a unique variant of a paired-comparison technique (Baron, 1996; Bowen, Martin, & Hunt, 2002; Johnson, Wood, & Blinkhorn, 1988). Data are obtained intra-personally; therefore, one stimulus can only be treated as a preference over another stimulus. Each item is a ranking of stimuli with no standard interval (Bowen et al., 2002; Cornwell & Dunlap, 1994).

Cornwell and Dunlap (1994) have indicated that because ipsative scores are intra-individual there is no way to establish meaningful differences among individuals. Individuals, therefore, cannot be ordered on a dimension that was measured ipsatively. The authors continue that while one can convert raw scores into rank orders or ipsative ranks, the absolute difference in the raw scores is lost in this process. Therefore, one is unable to derive a rank ordering of the stimuli from ipsative ranks and vice versa, which limits the ability to provide meaningful interpretation of scores compared to other individuals (Cornwell & Dunlap, 1994).

Because of these two main limitations, ipsative data are inappropriate for use in parametric analyses. As such, statistical analyses involving means, standard deviations, and correlations are likely to produce biased results that cannot be properly interpreted. It is also extremely difficult to separate typical scores on any given variable from deviant scores on the same variable between participants. Ipsative rank data should only be interpreted within subject and not between participants (Baron, 1996; Cattell, 1944; Closs, 1996; Cornwell & Dunlap, 1994; Johnson et al., 1988).

In order to discriminate deviance from typical responses, one must first have a normative reference group upon which to base future responses. Specifically, in order to understand deviant sexual interest, one must first establish normative sexual interest. As mentioned, inter-individual comparisons are not recommended and potentially harmful using ipsative data (Closs, 1996; Johnson et al., 1988).

Reference Group Comparison

The concerns about using ipsative data in determining "deviance" simply rest on how we define deviance - specifically deviance from what? When using intra-individual viewing time data, deviance can only suggest that one stimulus is more or less preferred than another within a

single subject. This does not make any references to whether this preference pattern is typical of others in the same population, nor could it predict future preference within the same subject.

By creating and utilizing norm-referenced data on sexual attraction, it is possible to eliminate the ambiguity of intra-individual comparison and allow for a comparison to an expected outcome. That is, instead of comparing one individual's preference of a stimulus to his or her preference on another stimulus, norm-referencing allows us to compare an individual's preference of stimulus to the expected preference of that stimulus for a particular sample of individuals. Using the Affinity 2.5 this can be done in two ways. First, we can compare a category specific viewing-time, Adult Female (ADF), to the expected viewing time for that category. Second, we can compare the overall profile pattern of all categories compared to the expected profile of all categories. It is prudent that when using norm-referencing, the individual's profile be compared to the profile drawn from a representative sample from which the individual can be assumed to belong.

In order to utilize norm-referencing, a reference sample must be selected, an expected pattern established, and temporal stability estimated. Harmon (2006) and Crosby (2007), using the Affinity 2.0, found a consistent pattern of viewing preference for college-aged undergraduate males and females who identified themselves as exclusively heterosexual. Boardman (2009) and Worsham (2009) replicated these findings using a similar sample and using the Affinity 2.5. Hansen (2012) demonstrated, with a third sample, a two week temporal stability of participants' viewing time profiles with reliability at 86% for males and 88% for females for the Affinity 2.5.

With an established and reliable norm-reference group of exclusively heterosexual college undergraduate males and females, the Affinity 2.5 may now be used to compare

individuals' profiles, assumed to belong to a similar population as the reference group, to the expected profile.

Current Study

It is apparent that when working with individuals whose sexual interests and behaviors are in question, clinicians need to be mindful of how they approach diagnosis, assessment, and treatment. By assessing sexual attraction based on norm-referenced data, clinicians may be able to more precisely identify differences or deviance within an individual that relate to risk for sexual offending. The purpose of this study is to examine the utility of the Affinity 2.5 to screen and diagnose individuals with possible deviant sexual interest. This study will use Fischer's Chi-square scoring procedure (see Data Analysis) to compare the viewing-time profiles of known sexual offenders to an expected profile based on several samples of exclusively heterosexual, non-pedophilic, males and females.

Research Questions

1. What percentage of convicted sexual offenders' will be correctly identified as having deviant sexual interest by the Affinity 2.5 using Fischer's Chi-square screening procedure?
2. What percentage of convicted sexual offenders' offense histories will be consonant with the diagnostic indicators on the Affinity 2.5 using Fischer's Chi-square scoring procedure?

Method

Participants

Data included in this study were collected from an outpatient clinic in Colorado. The individuals assessed had presented for psychological and sexual interest assessment relating to a

sexually based crime. State law required all individuals currently charged or previously convicted of a sexual crime to receive a psychosexual evaluation regardless of the nature of the current charges. The psychosexual evaluation was requested by the court or other interested parties to assist in sentencing. The data from 287 participants were included. All participants had been convicted of a sexual offense at some time in their life. Seventy-seven participants were evaluated specifically in relation to a criminal re-offense, only five of which were sexual offenses. The mean age of participants at the time of testing was 33.1 years, with a mean age at time of offense of 29.8 years. Participants completed the Affinity 2.5 as part of an individualized battery of assessments. Reference group data used for comparison were used from Hansen (2012) and included 63 male and 84 female undergraduate students who self-identified as exclusively heterosexual and non-pedophilic. The reference group participants ranged in age from 18 to 53 years old.

Procedures

Archival data were collected from a provider in an outpatient clinic in Colorado who performs psychological and psychosexual evaluations for the Colorado Department of Corrections. As part of the psychosexual evaluation, participants completed the Affinity 2.5. Affinity 2.5 results were paired with demographic data and de-identified to protect the identity of participants. Demographic data were collected from the assessment report and include gender, age at time of testing, age at time of offense, nature of offense, reason for current assessment, and known victim demographics – age, gender, and number of known victims.

Data were de-identified and coded. Analysis variables included Affinity 2.5 On-Task Latency (OTL) category totals, offender's age at time of offense, offender's age at assessment, and victim age and gender were categorized to match Affinity categories (e.g., juvenile female,

small child female). Offenses were categorized by the nature of offense into five groups: Incest, Child Pornography, Rape, Molestation, and Indecent Exposure. Two additional variables identified whether the participant had a non-sexual re-offense and if the participant had multiple victims.

For the purposes of this analysis categories were defined as follows. *Incest* was defined as any form of parent-child or sibling-sibling relationship such as father/daughter, step-father/step-daughter, brother/sister, or mother/son. Offenses where the victim and offender were related but not in a parent-child or sibling relationship were not categorized as incest (e.g., cousins, uncle/niece). This distinction was made on the assumption that parent-child and sibling relationships are characterized by imputed trust and a power imbalance (Lesniak, Rudman, Rector, & David Elkin, 2006) whereas this assumption is harder to make in extended family situations. *Child Pornography* included any offense where the offender had possession of nude or suggestive images of children and adolescents under the age of 18. This included possession of videos and images obtained surreptitiously, or text-based images from consenting minors, but did not include sending images of adults to minors. *Rape* was defined as any offense that included consensual but illegal or nonconsensual penetration by sexual organs or objects. This included vaginal, anal, and oral penetration, but did not include digital penetration alone, which was categorized as molestation.

Molestation was defined as any unwanted or consensual but illegal touch that did not include penetration, except digital penetration as mentioned above. This included fondling genitalia, breasts, anus, or any other body part in a sexually suggestive way. Molestation also incorporated masturbation of the victim, the offender, or both. *Indecent Exposure* included those

offenses where the offender exposed him or herself to others in a sexual way without any physical touch. This included public masturbation and exhibitionism.

Measure

The Affinity 2.5 is a computer-based assessment that purports to measure an individual's sexual interest by measuring viewing time. Participants look at 80 non-pornographic (fully clothed) pictures of males and females from four different age groups (adult, juvenile, pre-juvenile, and small child). It was originally designed to measure sexual interest in males with intellectual disabilities. It has since been licensed for use with adult male sex offenders with and without intellectual disabilities. Further, Affinity 2.5 has shown some utility in research and evaluation of adult and juvenile offenders of both sexes (Glasgow, 2003). The Affinity 2.5 creates participant profiles based on ranking and rating tasks.

Part one: Line drawing ranks. The ranking task requires participants to rank a series of line drawings according to sexual attractiveness and unattractiveness. The line drawings represent males and females in each of the four age groups. Participants select the type of person they find most sexually attractive from the line drawings of the eight categories. Participants then continue rating the line drawings in terms of sexual attractiveness until (s)he states there are no images left that (s)he finds sexually attractive. The participant then selects the type of person that they find most sexually unattractive from the line drawings. Again, the participant continues until all drawings have been ranked by either attractiveness or unattractiveness.

Part two: Ratings. After completion of the ranking task, participants are asked to rate the sexual attractiveness and unattractiveness of each of the 80 real life photographs of persons, on a 15-point scale. Each of the eight categories is represented in the rating task with 10 photographic images. Images are presented randomly for each participant. Images are rated

from “very sexually unattractive” (-7) to “neutral” (0), to “very sexually attractive” (+7). As the individual rates the images, Affinity 2.5 covertly tracks how long the participant views each slide, in milliseconds. Affinity 2.5 captures two measures of viewing-time. The first is On-Task Latency (OTL), which measures the time from when the slide first appears until the time the participant selects a final rating. The second measure is Post-Task Latency (PTL), which measures the amount of time from when the participant selects a final rating until (s)he changes to the next image.

The underlying assumption behind viewing-time measures of sexual attraction is that participants will spend little time evaluating a picture of someone (s)he is not sexually attracted to, and more time looking at pictures that meet at least minimum criteria for sexual attraction. This assumes that it takes additional time to categorize and rate the relative attractiveness of target persons and little time if the person is in a “non-attractive” category (Amoroso, Brown, Pruesse, Ware, & Pilkey, 1970; Crosby, 2007; Gress, 2005; Harmon, 2006; Harris, Rice, Quinsey, & Chaplin, 1996b; Israel & Strassberg, 2009; Quinsey, Rice, Harris, & Reid, 1993; Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996; Rosenzweig, 1942; Santtila et al., 2009; Worling, 2006; Wright & Adams, 1994; Zamansky, 1956).

The Affinity 2.5 has shown test-retest reliability (Hansen, 2012) of category viewing times for exclusively heterosexual, adult males and females and split-half reliability (Mokros, et al., 2013) of viewing times and self-report image ratings of males who are known pedophilic offenders, non-sexual offenders, and non-offenders.

Data Analysis

This study is an extension of previous studies examining the temporal stability of the Affinity 2.5 (Crosby, 2007; Hansen, 2012 ; Harmon, 2006). Using results from Hansen (2012) a

Chi-square goodness-of-fit approach was used to compare participant's viewing-time patterns to that of a sample of exclusively heterosexual, non-pedophilic, college aged males and females. Each offender's observed pattern was fitted against Hansen's (2012) norm-referenced expected viewing-time pattern to determine if the Affinity 2.5 is able to screen and diagnose sexual deviancy.

Fischer's Chi-Square Scoring Technique

Fischer and Meade (under review) has developed the Chi-square scoring technique wherein expected and observed data are compared based on the overall pattern of observed ipsative scores to an expected pattern of scores similar to how tests such as the MMPI-2 compare individual deviance on an individual scale to a norm-reference for that scale. Chi-square analyses have the following equation: $\chi^2_{(df)} = \sum_{i=1}^{j} \frac{(O_i - E_i)^2}{E_i}$,

wherein the differences of expected counts (E) within each category and observed counts (O) within each category are tested against chance. Fischer adapted this formula to:

$$\chi^2 = n * \sum_{j=1}^{J} \frac{(P_i - \pi_i)^2}{\pi_i}.$$

where *P* is the observed proportion of time viewed for each category and π is the expected proportion of time viewed for each category. Further, when using viewing time results, Fischer added *n* as a fair constant multiplier as a linear transformation to increase sensitivity and specificity of score variance. Specifically for the Affinity 2.5, Fischer suggests using the average total viewing time of a norm-reference group as the standard multiplier allows the researcher to create a fair comparison between those participants who were slower to respond to the instrument with those who were quicker to respond to the instrument (Fischer & Morgan, 2006; Fischer, Byrne, & Glasgow, 2007; Fischer & Meade, under review).

For Research Question 1, the Chi-square was computed from a conversion of raw On Task Latency (OTL) times per image stimulus and summed for each image category. Participant's OTL was divided by his or her Total Looking Time (TLT) to create an observed value. The same method was utilized to determine the expected value except the OTL used was the average summed OTL time for each image category divided by the average TLT obtained from Hansen (2012). Because the data is converted from raw data into percent data, the Chi-square value is then multiplied by a Fair Constant Multiplier (FCM) to transform the data back into raw form (Fischer & Meade, under review). The fair constant multiplier was selected based on a selectivity/specificity differential of a sample of reference group and offender data from Stephenson (2012). Statistically significant scores indicate significant deviance of the overall viewing-time pattern compared to the expected pattern.

Diagnostic statistics for Research Question 2 were computed using a Chi-square standardized residual. Residuals are the difference between a predicted score and the corresponding observed score (Howell, 2007). By standardizing residuals, we are able to compare residuals across categories with the formula: $\chi_r^2 = \frac{(O_j - E_j)}{\sqrt{E_j}}$ where *O* is the observed value and *E* is the expected value, and the residual is interpreted as a z-score (Elmore & Woehlke, 1997). To preserve data in the raw form, the residual formula was modified to convert percentage scores back to raw form as follows: $\chi_r^2 = \frac{n*(P_j - \pi_j)}{\sqrt{n*\pi_j}}$ using the same logic found in Fischer's Chi-square (Fischer & Meade, under review).

Results

A total of 287 potential participants were identified. From the 287 participants, those participants who were under age 18 were excluded. Forty-five participants with a sexual offense history only against adults were also excluded. An additional 37 participants were excluded

from analyses because their sexual crime was committed considerably in the past and the current rationale for assessment was a non-sexual re-offense. These exclusions resulted in a total of 158 participants. There were 155 males and 3 females with a mean age of 33.25 years with a mean age at time of offense of 31.91 years old.

Table 1

Participant Demographics and Offense Statistics

		Number	Percentage
Total		158	100.0
Males		155	98.1
Females		3	1.9
Offense Category			
Incest		45	28.5
Rape		57	36.1
Molestation		65	41.1
Child Pornography		30	19.0
Indecent Exposure		21	13.3
Single known victim		100	63.3
Multiple known victims		58	36.7
Has at least one PJ or SC victim		93	58.9
Identified Victim*:			
Juvenile Female	14-17 years	78	49.4
Juvenile Male	14-17 years	5	3.2
Pre-juvenile Female	9-13 years	46	29.1
Pre-juvenile Male	9-13 years	4	2.5
Small Child Female	0-8 years	35	22.2
Small Child Male	0-8 years	4	2.5

*excludes offenses of child pornography without an identified victim and includes multiple victims per participant

Of the 158 participants, 98 had a conviction in only one offense category, 54 had a conviction in two offense categories, and four had a conviction in three or more offense categories, and two participants had unknown offense category. Fifty-eight participants had more than one identified victim, whereas the majority of participants (n=100) had only one identified victim. Additional participant demographics are found in Table 1. Reference group

data from Hansen (2012) was used for comparison and included data from 84 females and 63 males ages 18 to 56.

When comparing the viewing-time profiles of convicted sexual offenders to that of an expected normative profile, Fischer's Chi-square method identified 69 (43.7%) of the 158 profiles as being statistically deviant from expectation. Of the 158 participants who offended against at least one protected population (i.e., juvenile, pre-juvenile or small child), 93 offended against pre-juvenile or small child victims. Fischer's Chi-square identified 40 (43.0%) of these individuals as being statistically deviant from expectation. Table 2 details the number of individuals identified with deviant sexual interest across other offense categories, and includes multiple offense categories per participant.

Table 2

Participants Identified as Deviant from Expectation, by Category

	Significant Chi-square	%	Significant Residual Only	%	Total n	Significant Chi-square or Residual	%
Identified as Deviant from Expectation	69	43.7	15	9.5	158	84	53.2
Males	67	43.2	15	9.7	155	82	52.9
Females	2	66.7	0	0.0	3	2	66.7
Incest	25	53.3	4	8.9	45	29	64.4
Rape	21	36.8	4	7.0	57	25	43.9
Molestation	34	52.3	4	6.2	65	38	58.5
Child Pornography	12	40	6	20.0	30	18	60.0
Indecent Exposure	7	33.3	1	4.8	21	8	38.1
Single known victim	49	49	8	8.0	100	57	57.0
Multiple known victims	20	33.9	7	11.9	59	27	45.8
Victim was a juvenile	35	42.2	7	11.9	83	42	50.6
Victims was a small child or pre-juvenile	40	43	8	11.9	93	48	51.6

The majority of participants in this sample had only one known victim (n=100), and had offended pre-juvenile or small child victims (n=93), rather than adolescent victims (n=65). The most common offense types were Molestation (n=65) and Rape (n=57). Nearly all offenses were against female victims, and nearly half of all female victims were juveniles (n=78, 49.1%).

Fischer's Chi-square method identified the greatest percentage of offenders whose offense category included incest (53.3%) or molestation (52.3%). Indecent exposure and having multiple known victims were identified as deviant the least, with 33.3 and 33.9%, respectively.

When assessing the diagnostic validity of Affinity 2.5 and Fischer's Chi-Square method, we find that of those participants who were identified as having deviant sexual interest and offended against a protected population, 36 participants (22.8%) also had a significant residual in a protected population category. That is, of the individuals who offended against a protected population and showed a significant Chi-square, 22.8% also had at least one category total viewing time that was more than two standard deviations above the mean for that category. Of those who offended against pre-juvenile or small child victims, 11 participants (11.8%) had a significant Chi-square and a significant residual in a pre-juvenile or small child category. Results detailed in Table 3 reflect the number of significant residuals including multiple significant residuals or offenses for a single participant.

Alternately, there were an additional 15 participants (9.5%) who offended against a protected population that had a significant residual in a protected category but did not have a significant Chi-square. Among those who offend against underage victims, using Fischer's Chi-square method on the Affinity 2.5, 84 participants (53.2%) were identified as deviant from expectation through either the overall Chi-square or residual.

When comparing offense history to the diagnostic category, 19 participants (12.0%) showed a match between their offense history and at least one significant residual category. That is, if a participant offended a pre-juvenile female victim, (s)he had a significant residual in at least the Pre-Juvenile Female (PJF) category.

Table 3

Fischer's Chi-Square and Residual Results, by Category

	Had significant					
	Residual		Chi-square & Residual		Consonant with Offense**	
Victim Category	n=	%*	n=	%*	n=	%*
Juvenile	33	20.9%	23	14.6%	15	9.5%
JUF	18	11.4%	11	7.0%	13	8.2%
JUM	15	9.5%	12	7.6%	2	1.3%
Pre-juvenile	14	8.9%	11	7.0%	4	2.5%
PJF	9	5.7%	7	4.4%	4	2.5%
PJM	5	3.2%	4	2.5%	0	0.0%
Small Child	7	4.4%	7	4.4%	1	0.6%
SCF	3	1.9%	3	1.9%	1	0.6%
SCM	4	2.5%	4	2.5%	0	0.0%

*Percentage of total N=158

** Had at least one significant residual

Affinity 2.5 and Fischer's Chi-square method were able to correctly identify 43.7% of individuals who had offended against children and adolescents, and 12.0% of these individuals demonstrated consonance between offense history and significant residuals. Given that all participants have at least one conviction for a sexual offense, the Affinity 2.5 identified less than half of these individuals as deviant from expectation. When considering the possibility of randomly assigning an individual into either an offender or non-offender category, we would

assume that the odds of correct assignment would be 50%. Since the screening identified fewer individuals as deviant than would be expected by chance, using Fischer's Chi-square approach is not a sufficient process for a recommendation of using the Affinity 2.5 as a screening or diagnostic tool in assessing deviant sexual interest.

Discussion

The results of this study indicate that 43.7% of convicted sexual offenders against children were identified as having a deviant sexual interest. Among those who offended younger children (pre-juvenile and small child), 43.0% were identified as having a deviant sexual interest. Fischer's Chi-square method is not a valid approach for screening for deviant sexual interest as it identifies deviant sexual interest no better than chance. When using the Affinity 2.5 and Fischer's Chi-square method as a diagnostic tool, it correctly identified only 22.8% of individuals as having a deviant residual in a protected population category. Further, of those with a significant Chi-square and a significant residual, only 12.0% of those individuals' significant residual category was consonant with their offense history. It seems that Affinity 2.5 and Fischer's Chi-square do not provide sufficient discrimination between those who sexually offend against children and adolescents and those who do not. These results align with that of Mokros et al. (2013), using Receiver Operating Characteristic (ROC) analysis, which suggested that use of Affinity 2.5 could discriminate between pedophilic offenders and non-pedophilic offenders, but that the results would include a high number of false positives.

Questions for Consideration

Using Fischer's Chi-square, the Affinity 2.5 had a low sensitivity for discriminating sexual offenders from non-offenders. There are several possible explanations for this finding. First, the reference group consisted primarily of undergraduate male college students aged 18 to

23, whereas the mean age for the offender group was 33.25 years old. The difference in mean age and emotional and sexual development between these two groups may indicate that the reference group is not an adequate group for comparison. What group, then, would be a proper reference group for sexual offenders? It is possible that a reference group comprised of older male participants may show greater discrimination between the reference group and the offender group? Perhaps a reference group of non-sexual offenders (e.g., violence, theft) rather than non-offenders would be a more precise reference group.

Second, the application of the premise that viewing-time is a measure of sexual interest may be faulty when applied as a screening tool. The Affinity 2.5 has been shown to be highly sensitive to subtle changes in total viewing time, (Fischer & Meade, under review) and perhaps this sensitivity is too great for normative application. When asked to rate the sexual attractiveness of an image the assumption is that individuals will spend more time evaluating images of persons that loosely fit or meet their opinion of a preferred sexual partner. It is possible that the Affinity 2.5 does not effectively capture this assumption. It is possible that the delay in viewing time is related to distractions in the testing environment, novelty or complexity of the image being viewed, or physical attractiveness unrelated to sexual behavior or preference (recognizing society's assumptions of attractiveness). It may also be possible that the environment or circumstance by which an individual is taking the Affinity 2.5 may influence the results. Individuals from the offender group may believe, to some degree that the Affinity 2.5 would provide some form of information about their propensity for acting out sexually. This priming may be a significant motivator for individuals to attend to images differently than a college student taking the Affinity for extra credit in a class.

Lastly, one of the greatest concerns in using the Affinity 2.5 and other sexual interest instruments is the problem of ipsative data and its inability to be normed across participants. In trying to get around this issue and create a profile pattern of viewing time, Fischer's Chi-square method inherently replicated this problem by turning raw scores into percentiles as a way to standardize the total looking time across participants. This process took raw data and created ipsative data that summed to 1.00. As mentioned above, ipsative data are such that data from one category directly influences the data of other categories. That is, if the proportion of viewing time in any one image category (e.g., ADF), is low then the proportion of viewing time in at least one other category will have to be high in order to sum to 1.00. It seems that unless, or until, the problem of ipsative data surrounding sexual interest measures is resolved, normative referencing may not be possible.

Study Limitations

There are many limitations of this study. First, there is a significant limitation in the reference group used for comparison. As mentioned above, the average age of the offender group was 33.25 years old while the reference group was predominantly under the age of 25. Other demographic information that was not collected may also influence the acceptability for comparison between groups including ethnicity, socioeconomic status, and education level. Without knowing if the two groups were consonant on these factors, it cannot be assumed that reference group obtained from Hansen (2012) is adequate for the offender group.

Second, because the offender data was archival, there is no way to guarantee similar administration and data collection practices. When consulting with the clinician, it was indicated that there were a few different common scenarios by which data was collected. First, individuals required a psychosexual assessment for a number of possible reasons and the individual

presented, of his/her own accord, to the clinician for assessment. Second, and more commonly, the individual had a court order for psychosexual assessment. In this circumstance individuals came to the clinic escorted by a parole officer or court official, or assessment may have been administered while the individual was incarcerated at the jail or prison. The differences in circumstances will likely have a significant impact on the individual completing the assessment affecting the administration and results.

Last, the individuals in the offender group were similar primarily in that they had all been convicted of a sexually based crime against a child or adolescent. The nature of the crime varied widely from viewing or creating child pornography and voyeurism, to the hands-on offenses of molestation or rape. Perhaps the Affinity 2.5 and Fischer's Chi-square method would have more sensitivity to a narrower population of offenders. If so, this begs the question of utility if screening is only successful for those individuals with particularly severe propensity for sexual offenses.

Future Directions

This study is a continuation, though not a conclusion, of a body of research for Affinity 2.5 and its uses as a screening or diagnostic tool for deviant sexual interest and preference. Reliability of a normative reference group for Affinity 2.5 has been established (Boardman, 2009; Hansen, 2012; Worsham, 2009) and a reference group expected profile has been tested against known offenders. While the results of this study do not show great promise for the Affinity 2.5 as a screening tool, other reference group categories should be established. These can include reference groups of varying ages, ethnicities, socioeconomic status, sexual orientation and any other group of individuals that could serve as an appropriate reference group.

With a greater array of reference group information, better discrimination between those who sexually offend and those who do not may be possible.

This study also tested Fischer's Chi-square as a statistical method for discriminating profile differences between offenders and non-offenders. Additional studies and tests with Fischer's Chi-square may be needed to ascertain its use in reference group comparison statistics. Additional statistical methods may also prove beneficial for screening and diagnostic applications for the Affinity 2.5.

References

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: Journal of Research and Treatment, 16*(3), 255-265.

Abel, G. G., Lawry, S. S., Karlstrom, E., & Osborn, C. A. (1994). Screening tests for pedophilia. *Criminal Justice and Behavior, 21*(1), 115-131.

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.*

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*(1), 49-56.

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, UT.

Bowen, C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis, 10*(3), 240-259.

Brown, M. (1979). Viewing time of pornography. *Journal of Psychology, 102*(1), 83.

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

Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology, 69*(1), 41-47.

Cornwell, J. M., & Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville & Willson (1991). *Journal of Occupational and Organizational Psychology, 67*(2), 89-100.

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

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.

Elmore, P. B., & Woehlke, P. L. (1997). *Basic statistics*. New York, NY: Longman.

Finkelhor, D., & Lewis, I. A. (1988). An epidemiologic approach to the study of child molestation. *Annals of the New York Academy of Sciences, 528*, 64-78.

Fischer, L., Byrne, P. M., & Glasgow, D. V. (2007). 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. (under review). Assessment of sexual interests: Reference-group solutions to problems of ipsative data. Unpublished manuscript.

Fischer, L., & Morgan, D. (2006). 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.*

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

Gannon, T. A., Rose, M. R., & Ward, T. (2008). A descriptive model of the offense process for female sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 20*(3), 352-374.

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*(2), 96-102.

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

Hansen, K. S. W. (2012). *Reliability and a measure of sexual interest: Examining the temporal stability of scores on affinity 2.5.* (Unpublished doctoral dissertation). Brigham Young University, Provo, UT.

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

Harmon, K. (2006). *The temporal stability of a normal heterosexual female response to affinity 2.0.* (Unpublished doctoral dissertation). Brigham Young University, Provo, UT.

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

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

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*(3), 167-184.

Howell, D. C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Thomson Wadsworth.

Israel, E., & Strassberg, D. S. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior, 38*(4), 551-558.

Johnson, C. E., Wood, R., & Blinkhorn, S. F. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology, 61*(2), 153-162.

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

Kinsey, A. C., Pomeroy, W. R., & Martin, C. E. (2003). Sexual behavior in the human male. *American Journal of Public Health, 93*(6), 894-898.

Lesniak, K. T., Rudman, W., Rector, M. B., & David Elkin, T. (2006). Psychological distress, stressful life events, and religiosity in younger African American adults. *Mental Health, Religion & Culture, 9*(1), 15-28.

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*(1), 93.

Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology (London, England: 1953), 92*, 239-255.

Mokros, A., Gebhard, M., Heinz, V., Marchall, R. W., Nitschke, J., Glasgow, D. V., Gress, C. L. Z., & Laws, R. (2013). Computerized assessment of pedophilic sexual interest through self-

report and viewing time: Reliability, validity, and classification accuracy of the Affinity Program. *Sexual Abuse: A Journal of Research and Treatment, 25*, 230-259.

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 S. M. Hudson (Ed.), *The juvenile sex offender.* (pp. 143-163). New York, NY US: Guilford Press.

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

Radcliffe, J. A. (1963). Some properties of ipsative score matrices and their relevance for some current interest tests. *Australian Journal of Psychology, 15*(1), 1-11.

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

Santtila, P., Mokros, A., Viljanen, K., Koivisto, M., Sandnabba, N. K., Zappalà, A., & Osterheider, M. (2009). Assessment of sexual interest using a choice reaction time task and priming: A feasibility study. *Legal & Criminological Psychology, 14*(1), 65-82.

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

Stephenson, H. (2012, October) Validity of Affinity 2.5 in an Incarcerated Population. *In: L. Fischer (Chair) Reliability, Falsification and Validity of Affinity 2.5. Symposium conducted at the meeting of the Association for the Treatment of Sexual Abusers Conference, Denver, CO.*

The National Center for Victims of Crime. (2010). 2010 national crime victims' rights week resource guide. Retrieved 04/21, 2010, from http://ovc.ncjrs.gov/ncvrw2010/pdf/6_StatisticalOverviews.pdf

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*(4), 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 Ph.D.). Brigham Young University, Provo, UT.

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*(4), 436.

Appendix: Literature Review

Assessment

Measuring sexual interest often used in judicial proceedings when an individual is accused or convicted of sexually assaulting another person. In these settings there are several methods of identifying deviant sexual interest for individuals who have been accused of a sexual crime. Deviance can be defined either through the Diagnostic and Statistical Manual (DSM) diagnoses, or through socially criminal behavior patterns, though each definition may have different implications. For example, there are many DSM paraphilic diagnoses while a diagnosis of a paraphilia does not necessarily mean an individual is breaking the law. However, many times when sexual behavior is criminal, it may also be diagnosable through the DSM. For the purposes of this paper, deviance will be restricted to those behaviors that are criminally deviant. The most common measures for sexual deviance include self-report, clinical interview, penile and vaginal plethysmography, and viewing-time measures (Quinsey, Rice, Harris, & Reid, 1993). Self-report and clinical interviews are subject to a number of response biases. Given the nature of the report and the face validity of the questions it is not difficult for respondents to give socially desirable responses instead of responding truthfully (Marshall, 1996; Quinsey et al., 1993). Self-report becomes less reliable when an individual is reporting for legal purposes, i.e., in response to criminal accusations or charges. These individuals are likely to be less honest in reporting true sexual feelings if they are deviant or socially questionable.

Plethysmography. Penile and vaginal plethysmography (PPG) is an objective measure of sexual interest, though it is highly intrusive. PPG is an instrument that measures genital tumescence, and is the most scientific method of assessing sexual interest (Quinsey & Chaplin, 1988). While PPG is one of the most researched and utilized assessment of sexual deviance,

several researchers have observed limitations with the reliability of the results (Adams, Motsinger, McAnulty, & Moore, 1992; McAnulty & Adams, 1991; Montano, 2006; Reyes, 2009). A major critique is related to the application of PPG. Many researchers have addressed that that assessment of genital tumescence is not the same as assessment of sexual arousal or attraction (Howes, 2003; Letourneau, 2002). PPG determines deviancy based on an individual's base rate and all observations are ipsative (or with-in individual) wherein deviance is determined by the individual, not by social convention or legal viewpoints. Specifically, if an individual demonstrates tumescence when viewing stimuli of young male children, but not to adult women the results suggest that he has greater interest in young children than to adult women. This is not to say that he is a pedophile, but simply that his measure of attraction to adult women is lower than that of young male children. There is no way to compare this to other individuals and suggest that this male is homosexual (socially defined) or a pedophile (legally defined) because there can be no extra-individual comparison.

A second area of concern is the potential for individuals taking the PPG to alter their results by intentionally suppressing their sexual responsiveness. Adams et al. (1992) found that when individuals were instructed to intentionally suppress their penile tumescence, both homosexual and heterosexual males were able to reduce their maximum tumescence by approximately 35%. Adams also found that when instructed to increase tumescence to a non-preferred stimulus; neither group was able to reliably enhance responsiveness.

Third, Marshall and Fernandez (2003) address the concern that administration of the PPG has not been standardized. Currently the PPG is administered in a number of settings, such as prisons, clinics, laboratories, and it is posited that the location of administration may alter responsiveness. Further, Lalumière and Quinsey (1994) (Lalumière & Quinsey, 1994)in a meta-

analysis of assessment of male rape offenders found that the images used for PPG pertaining to arousal to violent rape images, had significant differences in effect sizes depending upon which stimulus set was utilized.

Finally, there has been little evidence that PPG has test-retest reliability (Becker, Hunter, Goodwin, & Kaplan, 1992; Hunter, Becker, & Kaplan, 1995) and few studies employ repeated measures methodology. Reyes (2009) suggests that using repeated measure assessment of sexual interest may allow for a more complete analysis of an individual's sexual arousal.

Viewing-time measures. The most recent type of measure for sexual interest is a measure of viewing-time. Viewing-time (VT) instruments are used based on the components of the sexual arousal proposed by Singer (1984). He asserts that in determining sexual arousal, first an individual has an aesthetic response to a stimulus. This response leads the individuals to attend visually to the stimulus. This attention then leads to the second response: approach. During this phase, an individual desires to move closer, touch, be touched, or otherwise physically respond to the stimulus. The nature of the physiological response may produce any number of somatic responses including increased eye contact, and reducing the distance between self and object. The last phase of sexual arousal according to Singer is the genital response. This response includes genital tumescence as well as other autonomic responses such as increased heart rate and respiration. Singer suggests that reducing sexual attraction and responsiveness simply to genital responsiveness may be dehumanizing and misleading. By attending to the psychological and somatic responsiveness of an individual, we may be able to better understand human sexual attraction (Singer, 1984).

Viewing-time measures specifically address the aesthetic response of Singer's model with the assumption that individuals, when asked to rate the relative sexual attractiveness of an

individual from an image, will spend more time viewing the image if the object portrayed is comparable to the individual's ideal companion (Rosenzweig, 1942). When asked to rate the sexual attractiveness of images that an individual considers highly unattractive, the respondent is less likely to spend time cognitively processing the stimulus, and will instinctively respond (Macrae & Bodenhausen, 2000). When asked to rate images that are closer to the respondent's ideal attractiveness, the respondent must cognitively process the image in order to classify and rate the image (Macrae & Bodenhausen, 2000; Macrae & Bodenhausen, 2001). VT measures have been successful in discriminating between males and females with homosexual and heterosexual orientation; individuals with and without pedophilia; and individuals who experience high guilt around sexual topics versus those who do not (Harris, Rice, Quinsey, & Chaplin, 1996; Love, Sloan, & Schmidt, 1976; Quinsey et al., 1993; Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996; Rosenzweig, 1942).

Additionally, VT measure studies have addressed some of the limitations of PPG. First, studies have demonstrated that because of the covert nature of obtaining the data, VT measures are resistant to intentional faking. Abel, Jordan, Hand, Holland, and Phipps (2001) suggests that though the task seems simple enough to respond falsely to, the true measure behind viewing time is not related to the individual ratings of attractiveness. Instead, individuals have to identify which stimulus group (e.g., child, adult, paraphillic, control) each slide belongs to, as well as attending to the slides from each group a similar amount of time, which is quantitatively different than the cumulative time for other stimulus groups. Although it appears that VT measures should be resistant to intentional faking, to date, no study has demonstrated this with individuals known (as opposed to assumed) to be faking (Abel, et al., 2001).

Second, each VT measure has a standardized set of stimulus images. The more familiar measures: the Able Assessment of Sexual Interest and the Affinity 2.5, which use non-erotic images specifically targeting paraphilia and sexual attraction, respectively. For both measures, it is customary for participants to view all images, not just those pertaining to a presenting concern. Third, the Affinity 2.0 and 2.5 have demonstrated good test-retest reliability (between 76% and 88%) for both males and females (Hansen, 2012; Harmon, 2006).

Ipsative Data

Current objective measures of sexual interest utilize ipsative data. Data are ipsative when the scores sum to a constant (Cattell, 1944; Hicks, 1970; Radcliffe, 1963). Specifically, data are ipsative when the summed data for each person is the same. Cattell (1944) further describes ipsative data as a "designating [of] scale units relative to other measurements on the person himself." He argues that ipsative is a term to describe the process of obtaining and comparing data within an individual. This ipsative data must exist on some measurable continuum, and one cannot mix, within any individual, different continuums (interest, ability, character) unless each continuum co-exists on some greater common continuum (i.e. aptitude for success in a particular field) (Cattell, 1944). Ipsative data are obtained in several different ways. Most commonly, data are ipsative when participants ordinally rank items on some continuum. These data will sum to a constant as a high score on one stimulus automatically denotes a lower score on another stimulus (Closs, 1996). Data can also be ipsatized when using Likert scales and controlling for the total score of each participant across all participants. In doing so each data point represents a deviation from the average and the deviations will invariably sum to zero for all participants (Baron, 1996). Data are also ipsative when examining preference scores. Preference scores may

manifest as rank orders and do not indicate what is liked, they can only indicate what is preferred within a given stimulus set (Closs, 1996).

Viewing time and PPG are ipsative measures of sexual interest. Penile plethysmography measures sexual interest in terms of penile tumescence. Participants are fitted to the PPG and presented with sexual material. The assumption with PPG is that participants will exhibit greater tumescence for stimulus that he finds sexually appealing (preference), and decreased or no tumescence for stimulus that is not sexually appealing. For any given stimulus, there is a ceiling of possible scores at the maximum level of tumescence, and a floor when flaccid. In effect, tumescence, though measured continuously, has limitations in range and variability of scores. There is an artificial floor at maximum flaccidity (assumed to be baseline), and an artificial ceiling at maximum tumescence. While an individual can have a measure of tumescence anywhere within this range, these measurements will be different for each individual and there are practical limitations in one's ability to drastically change tumescence measure in a single iteration. When interpreting PPG, practitioners assess results in terms of either proportion of maximum tumescence per stimulus category, or deviations of tumescence (from baseline or maximum) per stimulus category.

Viewing time measures covertly measure sexual interest by assessing how much time a participant spends on any given stimulus, and then sums the time spent on each category. The Affinity 2.5 has eight categories of slides: Adult Male (ADM), Adult Female (ADF), Juvenile Male (JUM), Juvenile Female (JUF), Pre-Juvenile Male (PJM), Pre-Juvenile Female (PJF), Small Child Male (SCM), and Small Child Female (SCF). The Affinity 2.5 reports scores in two fashions: raw data - in milliseconds per slide - and mean ranks. Mean ranks are computed by assigning the rank of 80 (total number of slides) to the image with the longest viewing time.

Each slide is ordinally ranked from 80 to 1 from longest viewing time to shortest viewing time. Slides are then summed according to the eight categories and a mean rank is computed for each category (Glasgow, Osborne, & Croxen, 2003). Therefore, each participant's data sums to 3240 and each mean rank is constrained by the total score. Other viewing time measures use similar, proprietary, calculations in determining sexual preference.

Statistical Implications of Ipsative Data

When using ipsative data, there are a number of limitations in statistical analysis and interpretation. The first, and perhaps most significant limitation of ipsative data is one of non-independence. Statistical analyses require observations to be statistically independent or the correlation matrix is considered singular. Singularity occurs when the determinant of a square matrix equals zero, indicating that the matrix has no regular inverse (Kaw, 2008). Because the matrix is singular, ipsative data do not have an interpretable covariance matrix because of distortions within the covariance matrix (Baron, 1996; Cheung, 2004; Closs, 1996). When the covariance matrix is singular, standard forms of statistical analyses (e.g., structural equation modeling) are inappropriate as they require a nonsingular matrix. It has been proposed that by deleting one of the variables one can create a nonsingular matrix to use in analyses, but this leads to a negatively biased covariance of the remaining variables (Clemans, 1966; Cornwell & Dunlap, 1994; Johnson, Wood, & Blinkhorn, 1988). Interpretation of this correlation matrix is then limited as participants whose scores for all variables are at the bottom of a distribution would have a similar, if not identical, matrix as a participant whose scores for all variables are at the top end of a distribution (Baron, 1996). Closs (1996) illustrated this point using ipsative and non-ipsative data from a sample of 2808 school students. He demonstrated that the intercorrelations of non-ipsative data from these students yielded only one negative correlation

among all variables. When the intercorrelations were examined using the ipsative matrix, 73% of the variable correlations were negative. Closs argues that only one of the correlation matrices can be correct, and it is most likely that the negative correlations from the ipsative matrix are spurious (Closs, 1996).

A second limitation of ipsative data is that it is essentially an ordinal level of measurement and a unique variant of a paired-comparison technique (Baron, 1996; Bowen, Martin, & Hunt, 2002; Johnson et al., 1988). Because data are obtained intra-personally one stimulus can only be treated as a preference over the other stimuli. Each item is a ranking of stimuli with no standard interval (Bowen et al., 2002; Cornwell & Dunlap, 1994). Cornwell and Dunlap (1994) state that because ipsative scores are intra-individual there is no way to establish meaningful differences among individuals. Individuals, therefore, cannot be ordered on a dimension that was measured ipsatively. The authors continue that while one can convert raw scores into rank orders or ipsative ranks, the absolute difference in the raw scores is lost in this process. Therefore, we are unable to derive rank orders from ipsative ranks and vice versa, which limits the ability to provide meaningful interpretation of scores compared to other individuals (Cornwell & Dunlap, 1994).

Because of these two main limitations, ipsative data is inappropriate for use in parametric analyses. As such, statistical analyses involving means, standard deviations, and correlations are likely to produce biased results that cannot be properly interpreted. It is also extremely difficult to separate typical scores on any given variable from deviant scores on the same variable between participants. Ipsative rank data should only be interpreted within participants and not between participants (Baron, 1996; Cattell, 1944; Closs, 1996; Cornwell & Dunlap, 1994; Johnson et al., 1988).

Normative Reference

In order to assess typical responses from deviant responses, one must first have a normative reference group upon which to base future responses. Specifically, in order to understand deviant sexual interest, one must first establish normative sexual interest. As mentioned, inter-individual comparisons are not recommended and potentially harmful using ipsative data (Closs, 1996; Johnson et al., 1988). Several researchers have examined data to establish a normative reference profile for the Affinity 2.0 and 2.5, and each will be discussed in greater detail.

Harmon (2006) used Affinity 2.0 data from 146 exclusively heterosexual, non-pedophilic female college students. The age range of participants was 18 to 56 years, with a mean age of 21.6. Harmon sought to examine if a typical response profile existed for this population on the Affinity 2.0. Specifically, Harmon analyzed the data using eight different scoring techniques and compared each to the raw data to identify which techniques produced distortions and to what degree these distortions existed. The eight techniques included: category sums, category means, category medians, ipsative rank scores, category ranks, weighted ranks, standardized ipsative scores and standardized ipsatized on-task latency (OTL) raw scores. *Category sums, means and medians* took the raw OTL time for each of the seven category specific slides (ADM, ADF, JUM, JUF, PJM, PJF, SCM, and SCF) and found the corresponding statistic for each category. *Ipsative rank scores* were computed by summing the raw scores across all categories, subtracting the individual total from an arbitrary constant, dividing by the number categories (8), and finally subtracting the resulting quotient from each raw score. *Category ranks* were calculated wherein each category received an ordinal ranking (1-8) for each individual. *Weighted ranks* were calculated by assigning a rank to each of the 56 slides based on the raw OTL viewing time where

the slide viewed the longest received a rank of 56 and the slide viewed the shortest was assigned the rank of 1. Slide ranks were then averaged for each category and the resulting quotient was a weighted category rank with a minimum of 4 and a maximum of 53. *Standardized ipsative scores* were calculated by converting raw scores into t-scores, transforming them into ipsative scores and then standardizing those scores. *Standardized ipsatized OTL raw scores* were calculated exactly like standardized ipsative scores except that the first transformation into t-scores was omitted.

Using raw Affinity 2.0 scores, Harmon (2006) found a consistent pattern of viewing preference for female participants. There were four of the eight analyses techniques that did not produce significant distortion in the data. Raw category sums, means, medians and ipsative weighted ranks all produced similar patterns of overall response across participants. Harmon established for all four types of analyses that female participants viewed slides from longest viewed to shortest viewed as follows: ADM, JUM, ADF, JUF, PJM, PJF, SCF, and SCM. Ipsative, category rank, standardized ipsatized and standardized ipsatized OTL raw scores all resulted in significant distortion of data and were not considered further.

Crosby (2007) conducted a study similar to Harmon (2006) using 120 exclusively heterosexual, non-pedophilic, male college students. The age of participants ranged from 18 to 45 years, with a mean of 22.98. His analyses were identical to those proposed by Harmon, with the exception that he did not calculate the standardized ipsative scores. Unlike Harmon's results, Crosby noted a different ordering of categories when considering time spent viewing each category and analysis method. When utilizing means and ipsative weighted ranks, Crosby found that males viewed categories in descending order as follows: ADF, JUF, ADM, JUM, PJF, SCF, PJM, and SCM. When utilizing medians, he found the order changed slightly to: ADF, JUF,

ADM, PJF, SCF, JUM, PJM, and SCM, where JUM moved from the fourth rank to sixth. All other categories retained their rank ordering.

Both of these studies identified a temporally stable response profile for exclusively heterosexual, non-pedophilic males and females using the Affinity 2.0. While some minor changes in rank ordering of categories appeared at retest, the Chi-square Goodness of Fit tests for temporal stability of the overall pattern revealed no significant differences among the categories for either study.

In 2008, Glasgow released an update to the Affinity 2.0; Affinity 2.5. The new instrument included 42.8% more slides than the previous version. Instead of 7 images per category, the Affinity 2.5 includes 10 images per category. Although response profiles had been identified for 2.0, it was requisite to ascertain whether the increase of images would significantly alter the response profiles identified by Harmon (2006) and Crosby (2007).

Boardman (2009) set out to identify if exclusively heterosexual, non-pedophilic male college students' profiles on the Affinity 2.5 were significantly different than the profiles obtained from the Affinity 2.0. Boardman had a total of 54 male participants with ages ranging from 18 to 47, and an average age of 23.9. Boardman only examined category means following both Harmon's (2006) and Crosby's (2007) analyses that demonstrated little to no distortion in the data. Boardman found that males viewed categories in descending order as follows: ADF, JUF, PJF, SCF, JUM, ADM, PJM, and SCM. This ordering is essentially similar to that found by Crosby (2007) with the Affinity 2.0, but Boardman notes that ADF and JUF still have significantly higher proportions of viewing time than the remaining categories and the differences of proportion means for the remaining six categories represents a difference of .009

or less. When comparing the profile patterns for males, the differences are illustrated in Figure 1.

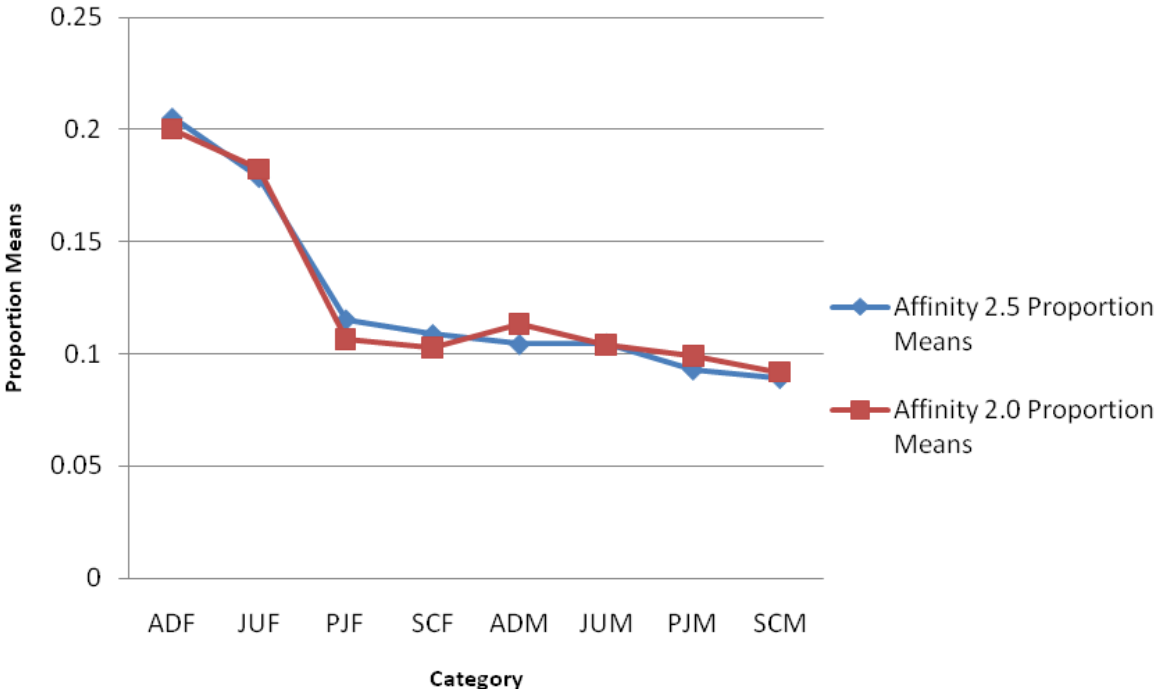

Figure 1. Replication of Boardman's (2009) comparison of Affinity 2.5 and Affinity 2.0 profiles.

Worsham (2009) wanted to discover if a response pattern existed for exclusively heterosexual, non-pedophilic college women on the Affinity 2.5 that was similar to the pattern established on the Affinity 2.0. Worsham had a sample of 63 female participants aged 18-26 with a mean age of 20.48. Similar to the study by Boardman (2009), Worsham examined the means for each stimulus category to create a viewing time profile. When comparing the differences in category orderings for the Affinity 2.0 and 2.5, Worsham found that the order changed slightly from ADM, JUM, ADF, JUF, PJM, PJF, SCF, and SCM on Affinity 2.0 to ADM, JUM, ADF, JUF, PJM, SCF, PJF, and SCM, where PJF moved from sixth position to seventh position. Like Boardman's study, however, Worsham found that ADM and JUM had proportion means that were significantly higher than the remaining categories and the differences

in the proportion means were less than .0196. When comparing the profile patterns for females on the Affinity 2.0 and 2.5, the differences can be seen in Figure 2.

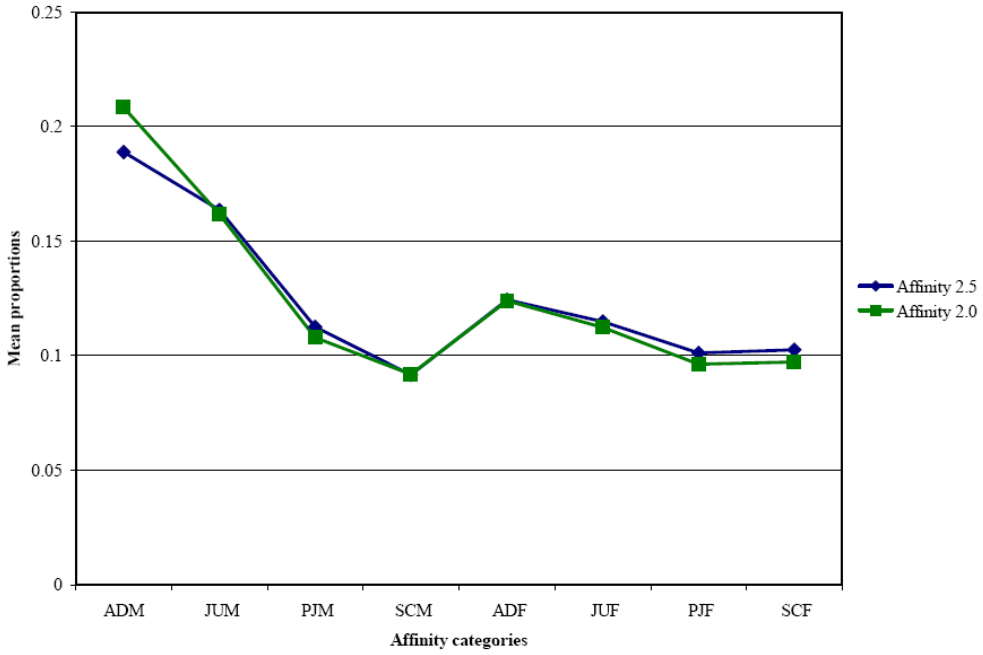

Figure 2. Replication of Worsham's (2009) comparison of Affinity 2.0 and Affinity 2.5 mean proportions.

Fischer's Chi-Square Scoring Technique

Fischer and Meade (under review) has developed the Chi-square scoring technique wherein expected and observed data are compared based on the overall pattern of observed ipsative scores to an expected pattern of scores similar to how tests such as the MMPI-2 compare individual deviance on an individual scale to a norm-reference for that scale. Chi-square analyses have the following equation:

$$\chi^2_{(df)} = \sum_{i=1}^{j} \frac{(O_i - E_i)^2}{E_i},$$

wherein the differences of expected counts (E) within each category and observed counts (O) within each category are tested against chance. Fischer adapted this formula to:

$$\chi^2 = n * \sum_{i=1}^{J} \frac{(P_i - \pi_i)^2}{\pi_i},$$

where *P* is the observed proportion of time viewed for each category and π is the expected proportion of time viewed for each category. Further, when using viewing time results, Fischer added n as a fair constant multiplier as a linear transformation to increase sensitivity and specificity of score variance. Specifically for the Affinity 2.5, Fischer suggests using the average total viewing time of a norm-reference group as the standard multiplier allows the researcher to create a fair comparison between those participants who were slower to respond to the instrument with those who were quicker to respond to the instrument (Fischer & Morgan, 2006; Fischer, Byrne, & Glasgow, 2007; Fischer & Meade, under review). Fischer purports that by utilizing the flexibility of Chi-square, one can use the totaled raw scores (instead of ipsatized scores) derived from intra-individual assessments, such as PPG and VT measures, to utilize the Chi-square distribution as an expected response in further Chi-square analyses.

Establishing Reliability

Affinity 2.0. Because normative reference profile pattern exists for exclusively heterosexual, non-pedophilic males and females, the next step required to use normative data is to establish temporal stability of these patterns. Both Harmon (2006) and Crosby (2007) assessed the temporal stability of response profiles by utilizing a two week delay from test to re-test administration of the Affinity 2.0. Harmon assessed temporal stability using Pearson Product Moment Correlations (PPMC) and Chi-square goodness-of-fit analyses. Harmon used PPMC to assess the correlation of category means from test administration to re-test administration to determine if each category mean was stable across test administrations.

Harmon found that all categories had correlations that were significant at the .01 level suggesting significant correlation of scores from test to re-test, with ADM and PJF having moderately strong correlations, and JUM, PJM, SCM, ADF, JUF, SCF having only mildly strong correlations.

Using the Chi-square analysis the mean ipsative weighted rank for each category (mean category rank computed through weighted stimulus ranks) on the test administration of Affinity 2.0 was set as the expected (E) and the re-test administration mean ipsative weighted ranks served as the observed (O) score with $\chi^2 = \frac{|O\text{-}E|^2}{E}$. Harmon found no significant deviation in observed scores compared to expected scores with $(\chi^2(7) = 0.29, p > .05)$, as well as finding that none of the eight categories yielded a residual greater than the ± 1.96 level. These results demonstrate that the overall category test proportions were similar to re-test category proportions and any difference was not significant and that the individual categories yielded similar proportions of time viewed from test to re-test, suggesting that the response profile for exclusively heterosexual, non-pedophilic female college students has temporal stability and that Affinity 2.5 is a reliable measure for assessing sexual interest for this population.

Crosby (2007) also conducted analyses using PPMC and Chi-square goodness-of-fit analyses to assess the temporal stability of the response profile for exclusively heterosexual, non-pedophilic male college students. Crosby's results were similar to those found by Harmon (2006). Crosby found that the PPMC were significant at the 0.01 level for four categories: ADF, ADM, PJF, and JUM. ADF and ADM held moderate strength correlations, PJF and JUM had mildly strong correlations and JUF, SCF, PJM, and SCM did not produce significant correlations. However, when Crosby analyzed the temporal stability through Chi-square analyses, he found no significant differences in test and re-test profile weighted ranks $(\chi^2(7) =$

$0.219, p > .05)$ and none of the standardized residuals surpassed a ± 1.96 significance level. His results also suggest that the expected profile for this population of male respondents does demonstrate temporal stability using the Affinity 2.0.

As previously discussed, Boardman (2009) and Worsham (2009) established an expected response profile for Affinity 2.5. The authors also reanalyzed the data from Crosby and Harmon, respectively, using a Chi-square goodness-of-fit analysis wherein the authors tested the fit of the observed retest overall response pattern to the expected test overall response pattern using the Fischer's Chi-Square formula:

$$\chi^2 = n * \sum_{i=1}^{J} \frac{(P_i - \pi_i)^2}{\pi_i},$$

instead of analyzing the individual category stability (Boardman, 2009; Worsham, 2009). Instead of focusing on the stability of categories across participants from test to re-test, Boardman and Worsham assessed the stability of an individual's response pattern from test to re-test. Both authors found the Chi-square coefficient by multiplying the sums of differences between the expected - test - proportion (π) and the observed - re-test - proportion (P) by a constant (n) to increase sensitivity and specificity. Boardman found that 76.6% of male participants demonstrated temporal stability of the overall profile from test to re-test administrations of the Affinity 2.0. Worsham similarly found that 86% of female participants demonstrated temporal stability of the overall profile from test to re-test administrations of the Affinity 2.0.

Affinity 2.5. With the introduction of Affinity 2.5, new reliability assessments were required to see if the temporal stability existed for Affinity 2.5, similar to that found on Affinity 2.0. Hansen (2012) assessed whether the increase in number of images per category would alter the temporal stability of the individual Affinity 2.5 profile compared to the individual Affinity

2.0 profile. Hansen had a sample of 67 male and 90 female exclusively heterosexual, non-pedophilic college students. Hansen found that the category proportion means using the Affinity 2.5 were very similar to those found on previous studies of Affinity 2.5 for both males and females (see Figure 3).

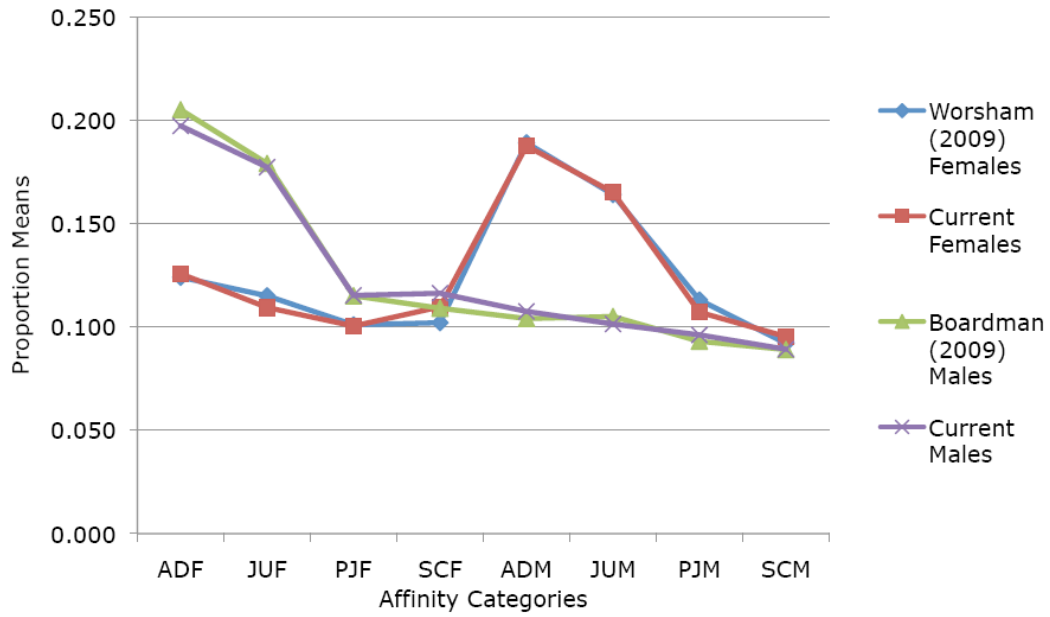

Figure 3. Reproduction of Hansen's (2012) proportion means for studies using Affinity 2.5.

Hansen tested individual temporal stability using Fischer's Chi-square procedure and found that 86% of male and 88% of female participants demonstrated temporal stability in Affinity 2.5 viewing time profiles. These results replicate those found by Boardman (2009) and Worsham (2009) suggesting that Affinity 2.5 demonstrates reliability for participant's overall viewing-time response patterns.

Reference Group Comparison

The concerns about using ipsative data in determining "deviance" simply rest on how we define deviance – specifically, deviance from what? By creating and utilizing norm-referenced

data on sexual attraction, it is possible to eliminate the ambiguity of this question. Instead of comparing VT profiles of an individual across categories, norm-referencing would compare an individual's entire VT profile with an expected profile drawn from a similar population.

Using the Affinity 2.5, Hansen (2012), Worsham (2009), and Boardman (2009), have created norm-reference data using exclusively heterosexual, non-pedophilic college age males and females. This approach created a curve of expected viewing-time profiles for this population. Hansen (2012) has also demonstrated the temporal stability of this profile.

The Affinity 2.5 may prove to be a beneficial screening and diagnostic tool by fitting an individual's viewing-time curve to the viewing-time curve of the norm-referenced sample using Fischer's Chi-Square approach. If there is not a significant deviance, it can be assumed that the individual does not have deviant sexual interest. The Chi-square approach may allow the Affinity 2.5 to function as a screening tool when there is a significant difference in the two curves, and further serve as a diagnostic tool by identifying any significant category-based deviance (i.e., significant deviance in one or more category, instead of overall deviance).

References

Abel, G. G., Jordan, A., Hand, C. G., Holland, L. A., & Phipps, A. (2001). Classification models of child molesters utilizing the Abel Assessment for Sexual Interest. *Child Abuse & Neglect, 25*(5), 703-718.

Adams, H. E., Motsinger, P., McAnulty, R. D., & Moore, A. L. (1992). Voluntary control of penile tumescence among homosexual and heterosexual subjects. *Archives of Sexual Behavior, 21*(1), 17-31.

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology, 69*(1), 49-56.

Becker, J. V., Hunter, J. A., Goodwin, D., & Kaplan, M. S. (1992). Test-retest reliability of audio-taped phallometric stimuli with adolescent sexual offenders. *Annals of Sex Research, 5*(1), 45-51.

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, UT.

Bowen, C., Martin, B. A., & Hunt, S. T. (2002). A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires. *International Journal of Organizational Analysis, 10*(3), 240-259.

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

Cheung, M. W. L. (2004). A direct estimation method on analyzing ipsative data with Chan and Bentler's (1993) method. *Structural Equation Modeling, 11*(2), 217-243.

Clemans, W. V. (1966). *An analytical and empirical examination of some properties of ipsative measures.* (14th ed.). Richmond, VA: The William Byrd Press, Inc.

Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology, 69*(1), 41-47.

Cornwell, J. M., & Dunlap, W. P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville & Willson (1991). *Journal of Occupational and Organizational Psychology, 67*(2), 89-100.

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

Fischer, L., Byrne, P. M., & Glasgow, D. V. (2007). 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. (under review). Assessment of sexual interests: Reference-group solutions to problems of ipsative data. Unpublished manuscript.

Fischer, L., & Morgan, D. (2006). 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.* Chicago, IL.

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*(2), 96-102.

Hansen, K. S. W. (2012). *Reliability and a measure of sexual interest: Examining the temporal stability of scores on affinity 2.5.* (Unpublished doctoral dissertation). Brigham Young University, Provo, UT.

Harmon, K. (2006). *The temporal stability of a normal heterosexual female response to affinity 2.0.* (Unpublished doctoral dissertation). Brigham Young University, Provo, UT.

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

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. *Psychological Bulletin, 74*(3), 167-184.

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

Hunter, J. A., Becker, J. V., & Kaplan, M. S. (1995). The adolescent sexual interest card sort: Test-retest reliability and concurrent validity in relation to phallometric assessment. *Archives of Sexual Behavior, 24*(5), 555-561.

Johnson, C. E., Wood, R., & Blinkhorn, S. F. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology, 61*(2), 153-162.

Kaw, A. K. (2008). *Introduction to matrix algebra.* (2nd ed.). Online Text: Autar Kaw.

Lalumière, M. L., & Quinsey, V. L. (1994). The discriminability of rapists from non-sex offenders using phallometric measures: A meta-analysis. *Criminal Justice and Behavior, 21*(1), 150-175.

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.

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*(1), 93.

Macrae, C. N., & Bodenhausen, G. V. (2001). Social cognition: Categorical person perception. *British Journal of Psychology (London, England: 1953), 92*, 239-255.

.Marshall, W. L. (1996). Assessment, treatment, and theorizing about sex offenders: Developments during the past twenty years and future directions. *Criminal Justice and Behavior, 23*(1), 162-199.

Marshall, W. L., Fernandez, Y. M., Serran, G. A., Mulloy, R., Thornton, D., Mann, R. E., & Anderson, D. (2003). Process variables in the treatment of sexual offenders: A review of the relevant literature. *Aggression and violent behavior*, *8*(2), 205-234

McAnulty, R. F., & Adams, H. E. (1991). Voluntary control of penile tumescence: Effects of an incentive and a signal detection task. *Journal of Sex Research, 28*(4), 557-578.

Montano, D. E. F. (2006). *An investigation of response interference in penile plethysmography*. ProQuest Information & Learning). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66*(11).

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., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology & Sociobiology, 17*(5), 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 S. M. Hudson (Ed.), *The juvenile sex offender.* (pp. 143-163). New York, NY US: Guilford Press.

Radcliffe, J. A. (1963). Some properties of ipsative score matrices and their relevance for some current interest tests. *Australian Journal of Psychology, 15*(1), 1-11.

Reyes, J. R. (2009). *Assessment of sex offenders with developmental disabilities.* ProQuest Information & Learning). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 69*(10).

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

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

Worsham, M. (2009). *Creating an expected profile for affinity 2.5 from a sample of non-pedophilic, exclusively heterosexual, college-age females.* (Unpublished Ph.D.). Brigham Young University, Provo, UT.