2008-07-10

Males' Ipsative Score Distortion on Affinity 2.0

Jeffrey Brian Madsen
Brigham Young University - Provo

MALES' IPSATIVE SCORE DISTORTION ON AFFINITY 2.0

by

Jeffrey B. Madsen

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Educational Specialist

Department of Counseling Psychology and Special Education

Brigham Young University

May 2008

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Jeffrey B. Madsen

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

_____	_____
Date	Lane Fischer, Chair
_____	_____
Date	Ellie Young
_____	_____
Date	Rachel Crook Lyon

BRIGHAM YOUNG UNIVERSITY

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

_____
Date

_____
Lane Fischer
Chair, Graduate Committee

Accepted for the Department

_____
Ellie Young
Area Coordinator
Counseling Psychology & Special Education

Accepted for the College

_____
Barbara Culatta
David O. McKay School of Education

ABSTRACT

MALES' IPSATIVE SCORE DISTORTION ON AFFINITY 2.0

Jeffrey B. Madsen

Department of Counseling Psychology and Special Education

Educational Specialist

This study investigated the frequency of distortion that occurs when raw score patterns of *Affinity 2.0*, a viewing time measure designed to assess sexual interest, are converted to ipsative scores. Eighty-eight percent of a sample of ninety-nine non-pedophilic, exclusively heterosexual males' profiles sustained some degree of distortion. The study also applied Brown's (2005) approach to predicting distortion with this sample of males' responses. Brown's techniques were largely ineffective in predicting males' distortion scores.

ACKNOWLEDGEMENTS

I would like to thank my wife, Jordan, for her love, encouragement, patience, and most importantly, support. Graduate studies and raising small children are not always complimentary. Over the last few years, there has been many times where she picked up my end of the load in addition to hers without complaining. She also has a way of listening and validating while encouraging without mixing words like no other. She is my best friend, companion, and confidant.

I would also like to thank my parents for their support throughout my education. They have been a great help, both emotionally and financially. I would not have been able to pursue my graduate degree without their assistance and concern.

Lastly, I wish to thank my thesis chair, Dr. Lane Fischer, for his guidance, time, and patience. He has always understood my circumstances and shown concern for my studies, family and personal life. Not only has our relationship been influential to me educationally, but personally and professionally as well.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The assessment of sexual interest can contribute significantly to the treatment of sex offenders. Assessment aids in the identification of preferred sexual stimuli and targets of deviant sexual behavior. Information obtained about the preferred stimuli can aid in the development of interventions and treatment plans aimed at eluding potential high-risk situations, and help to deter the commencement of an offense chain leading to perpetration by sexual offenders (Fischer, 2000).

The field of assessing deviant sexual interest (e.g., pedophilia, voyeurism, exhibitionism) is both evolving and complex. Although penile plethysmography (PPG) has been one of the most frequently used methods for evaluating deviant sexual interest, this procedure has serious limitations. For example, PPG participants are exposed to sexually explicit images, an unethical procedure in some cases. PPG is also a labor intensive, expensive, and highly invasive procedure (Laws & Gress, 2004), but perhaps the most serious deficiency in PPG is the lack of standardization. In 2004, Laws and Gress reported that there was not a standardized procedure for PPG administration, scoring, and interpretation of data—thus raising serious questions about the reliability and validity of the results. However, more recently, the *Monarch 21* PPG system has been standardized, but reliability studies are as yet unpublished (Cloyd, 2006).

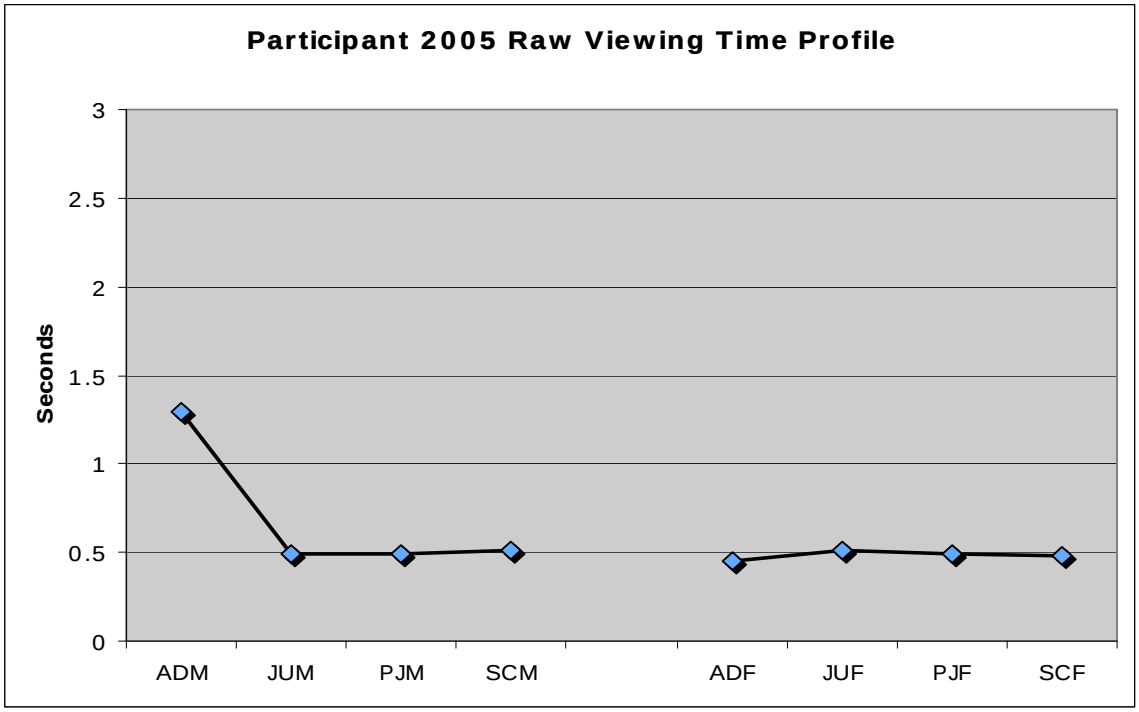
As a response to PPG limitations, researchers have tried to develop alternative methods for assessing deviant sexual interest. Viewing time (VT) assessment is one such method. Research indicates that people view preferred sexual stimuli longer than other non-preferred stimuli (Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996; Wright &

Adams, 1994; Zamansky, 1956). Therefore, in VT assessment, the participant is exposed to images of various potential sexual stimuli while VT is implicitly measured.

Two VT assessments are commercially available—*the Abel Assessment for Sexual Interest (AASI)* and *Affinity*. There appear to be serious questions about the reliability and validity of the *AASI* (Fischer, 2000; Fischer & Smith, 1999; Smith & Fischer, 1999). While there is some evidence for the reliability and validity of *Affinity* (Glasgow, 2003; Glasgow, Croxen, & Osborne, 2003), this evidence still needs further replication and support. A potential problem with both of these VT assessments is that clinicians continue to use their ipsative scores in the absence of norm-referenced scores. This can be problematic if the ipsative scores distort the true underlying picture.

Statement of Problem

Attempting to further examine the reliability of the *Affinity* assessment—in order to demonstrate the need for interpretation and decision making based on norm-referenced data—researchers at Brigham Young University recently gathered data on samples of non-offending, exclusively heterosexual males and females. While doing so, they noticed in some cases that the conversion from raw viewing times to ipsative scores resulted in distortion of the patterns. The ipsative score profile results in a pattern of sexual attraction differing from that of the raw viewing time profile. Figure 1 illustrates the distortion that occurred in the conversion from raw viewing time to ipsative scores for one such female *Affinity* participant:



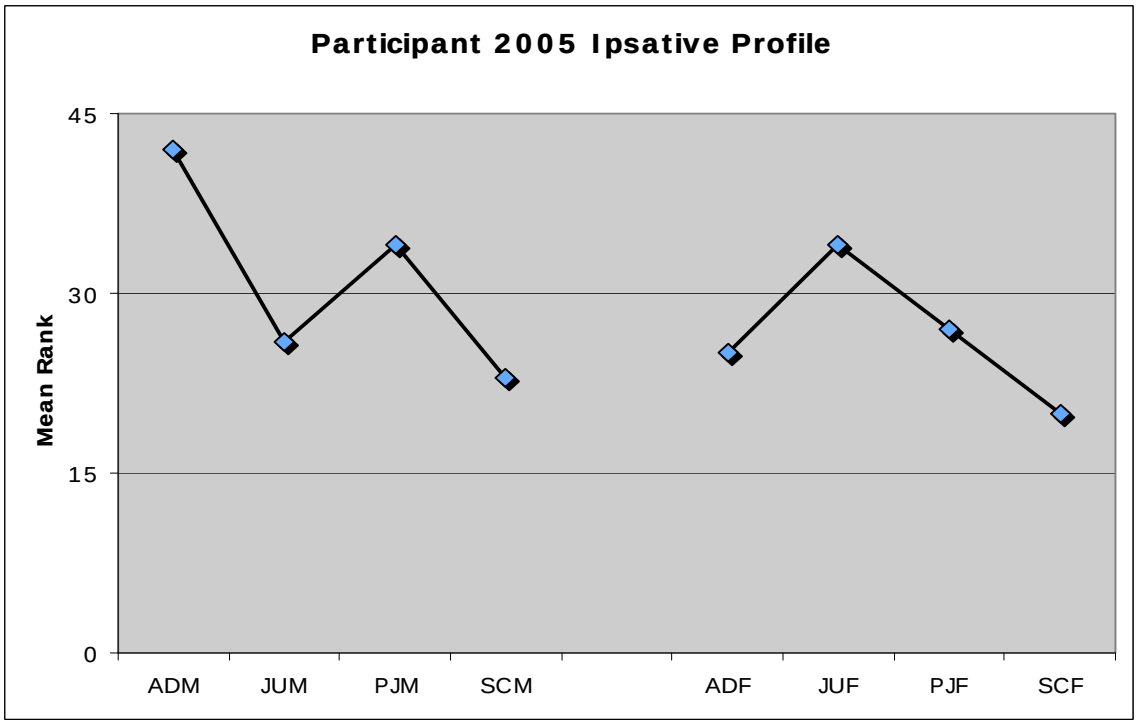


Figure 1. Raw score and ipsative conversion.

Being able to warn clinicians of the potential invalidity of a profile would decrease the probability of drawing erroneous conclusions regarding an examinee, and possibly indicate cases where the examinee was trying to obscure their true sexual preferences.

Brown (2005) developed a method to measure distortion objectively in the conversion of *Affinity* raw scores among a sample of 146 non-offending, exclusively heterosexual females. He determined that 86% of all profiles incurred distortion. He formulated and tested several hypotheses as to variables that might be related to distortion. The relationships were examined using Pearson Product Moment Correlations. Several statistically significant, but weak, correlations were found. An interaction effect was calculated for four of the variables, and was found to have a moderately strong correlation with distortion ($r = .530$, $p < .01$). Brown developed an algorithm for identifying distortion using the interaction effect. He tested several cut-off scores and the most effective cut-off correctly identified 42.9% of the significantly distorted profiles. It is unclear whether Brown's analytical technique will function similarly with males' responses to *Affinity 2.0*.

Statement of Purpose

The purpose of this study is to test Brown's (2005) model among the time-one sample of 99 non-pedophilic, exclusively heterosexual males and their responses to *Affinity 2.0* – an instrument that purports to assess sexual interest by surreptitiously measuring VT of non-pornographic images (Crosby, 2008). All profiles will be objectively measured and examined to identify distortion using Brown's analytical technique. The same variables that Brown hypothesized would predict distortion and their

interaction effects will be applied to the male sample. Implications and limitations of the results will be discussed, and directions for future research will be provided.

REVIEW OF LITERATURE

Relapse prevention posits that there is a likely course of offenses that sexual offenders follow leading to perpetration. Identifying deviant sexual interests (e.g., pedophilia, voyeurism, exhibitionism) of a sexual offender becomes critical in the development of treatment plans. Deviant sexual interests are assumed to be similar to lifelong addictions that must be handled through adaptive coping behavior. Abstinence from sexual deviancy is the goal. If the course of events in the relapse path can be made apparent to the offender, then they can be taught alternative patterns of thoughts and behaviors that will prevent relapse. Likely victim groups can be identified using the knowledge of deviant sexual interest. Potential triggers of the offense course, such as stimuli and environments, can be identified, and the offender's environment can be structured in such a way that exposure to these triggers is minimized (Fischer, 2000).

Ward et al. (1995) proposed the idea of an offense chain, or a set of stages, which sexual offenders progress through leading to perpetration. These stages include distal planning, contact with potential victims, proximal planning, offending, cognitive restructuring, and future resolution. Theoretically, it is possible to break this offense chain in the initial stages by arranging the environment of a sexual offender in such a way that the offender is able to avoid potentially high-risk situations before even offending.

Methods for Evaluating Deviant Sexual Interest

Although there is no standardized method to assess sexual preference, it generally involves self report, analysis of records, and clinical interviews (Laws, 1989; Marshall, 1996). Self reports and clinical interviews "are somewhat subjective and may be compromised by dissimulation" (Fischer, 2000, p.304). Therefore, other assessment tools

that are objective and less vulnerable to deceitful patterns of responding are needed. Two such methods currently in use are penile plethysmography (PPG) and viewing time (VT) assessment.

Penile plethysmography (PPG). Singer (1984) hypothesized that male sexual arousal included three distinct stages rather than a single global construct. Singer's trichotomy of sexual arousal include a) the aesthetic response—sustained attention to the given stimuli; attraction, b) the approach response—movement toward the stimuli; interest, and c) the genital response; arousal. Singer presumes that, although all three stages can be experienced independently, they do interact with one another. Singer states there is general agreement that genital changes are the most reliable and convenient to measure. However, operationalizing arousal, primarily in genital terms, may be dehumanizing. PPG measures sexual interest at the final stage – sexual arousal/penile engorgement. In PPG, a measurement device is attached to the individual's penis. The individual is then exposed to visual and/or audio depictions of various sexual objects or behaviors, while penile engorgement is recorded (Fischer, 2000).

Although PPG is more objective and less vulnerable to dissimulation than clinical interviews, this assessment procedure has many different drawbacks. PPG exposes participants to sexually explicit images, which may not be ethical with some individuals—most notably adolescents. Fischer (2000) stated that exposure to sexually deviant imagery, "may inadvertently introduce, stimulate, or ratify [deviant] behavior." An additional ethical concern with PPG is its highly invasive nature. PPG is also labor intensive, uncomfortable, requires expensive equipment, and can only be used with males (Laws & Gress, 2004).

In addition to the ethical concerns with PPG, another critical drawback exists. There has been limited standardized procedure for the administration, scoring, or interpretation of data; thus, seriously limiting the value of obtained data (Laws & Gress, 2004). Without standardization of administration, scoring, and interpretation, it is impossible to compare individuals to others around sexually deviant interests. It is also impossible to establish the reliability of PPG without standardization. The *Monarch 21* PPG system has recently been introduced and includes standardized administration, stimuli, scoring, and training (Cloyd, 2006). However, reliability studies of the *Monarch 21* have yet to be published. Since validity is dependent upon reliability, validity is likewise not established.

Viewing time. The development of several assessment methods has come as a response to the need for alternatives to PPG. One such alternative assessment method measures sexual interest at stage one of Singer's (1984) model – sustained attention to preferred sexual stimuli. Two commercially available tests, the *Abel Assessment for Sexual Interest (AASI)* and *Affinity 2.0*, measure the time spent viewing potential sexual stimuli in order to assess sexual interest. Before examining the *AASI* and the *Affinity*, it is salient to review and highlight that which is known about sustained visual attention as a predictor of sexual preference.

Rosenzweig (1942) created an instrument he called the "photoscope" to objectively evaluate the changes in sexual interest, due to the effects of hormone therapy, of 20 inpatients diagnosed with schizophrenia. The "photoscope" consisted of a set of sexual and non-sexual pictures, on a rolodex-type device that the subject was allowed to examine at leisure and without supervision. The participants were able to control their

time spent viewing each slide by deciding when to flip to the next picture. An observer behind one-way glass recorded the viewing time for each slide. The 20 participants were divided into two equal groups of high or low sexual interest based on staff observations of overt sexual behavior. The average viewing time of the non-sexual photographs for both groups was 19 seconds, whereas the average viewing time of the sexual photographs differed significantly between the high and low groups. The average viewing time of the sexual photographs for the high sexual interest group was 40 seconds, while the average viewing time of these same slides for the low sexual interest group was 13 seconds.

Zamansky (1956) used a similar, more sophisticated method to examine whether VT could be used to discriminate homosexual males from heterosexual males. He observed the eye movements of 20 homosexual and 20 heterosexual males viewing paired slides containing male, female, or neutral content. As with Rosenzweig (1942), the results supported VT as an indicator of sexual preferences. In the paired slides of males and females, the group of homosexual males viewed the pictures of men an average of 3.55 seconds longer than the group of heterosexual males, and the heterosexual males group viewed the pictures of women 2.37 seconds longer than the homosexual males group. Zamansky concluded that sexual preference "will manifest itself in the pattern of an individual's visual fixations, if these fixations can be measured without his awareness" (1956, p.446).

Ware, Brown, Amoroso, Pilkey, and Pruesse (1972) and Brown, Amoroso, Ware, Pruesse, and Pilkey (1973) measured sustained visual attention and investigated the meaning of pornographic stimuli through semantic differential ratings. Forty male college students participated in this study by individually controlling a slide projector, viewing a

set of 15 slides, and rating these slides along several dimensions – evaluative, activity, and potency factors. The 15 slides varied in sexual explicitness that ranged from a dressed couple entering a room to a nude female on a bed with her genitals exposed. Ware et al. (1972) discussed the results of all forty participants while Brown et al. (1973) reported the results of the same group divided into the two conditions of being alone or observed by others while viewing the slides. The results revealed that the higher the sexual explicitness and activity (i.e., non-stimulating vs. stimulating, passive vs. active) of the slide, the longer the VT (r=.66 & .93, respectively). In addition, Brown et al. found that VT was generally more variable and longer when subjects were alone than when in the presence of others.

Quinsey, Rice, Harris, and Reid (1993) also measured VT responses of heterosexual male and females to nude, non-explicit slides that varied in age and gender. The slides included nude photographs of male and female children, adolescents, and adults. The slides were not considered to be sexually explicit because none of the models were photographed in flirtatious poses. The authors found little variation in the participants' VT across the different ages and genders of the stimuli with the exception that the participants observed the slides of adults and pubescents of the preferred gender longer than all categories of the non-preferred gender.

Wright and Adams (1994) measured the sustained visual attention of 20 adult subjects in each of the four groups (heterosexual males, heterosexual females, homosexual males, and homosexual females) using neutral stimuli and sexually explicit slides from *Playgirl* and *Playboy* magazines to discriminate sexual preference. They hypothesized that VT would be considerably higher to slides of the preferred sexual

object compared to those of the non-preferred sexual object. As hypothesized, there were notable differences in VT between the four groups, with greater VT to the slides of the preferred sexual partners.

Harris, Rice, Quinsey, and Chaplin (1996) compared VT, sexual attractiveness ratings, and penile tumescence of male child molesters and normal heterosexual males. A total of 25 paid volunteers from the community who reported being heterosexual and 26 child molesters were shown neutral and nude slides of male and female adults, pubescents, and adults. The non-offenders' VT profiles, sexual attractiveness ratings, and phallometric responses were very much alike while the molesters' phallometric responses did not correspond to their sexual attractiveness ratings. The molesters' average VT was quite low (1.87 seconds per slide) and showed restricted variability while the non-offenders' average VT was 3.25 seconds per slide and showed variability across the stimulus categories. Harris et al. concluded that "VT can serve as an unobtrusive measure of males' sexual interests . . . [and] that the present results also show the promise of viewing time measures for screening men to work with vulnerable children" (p. 394).

Quinsey, Ketsetzis, Earls, and Karamanoukian (1996) conducted two studies using normal heterosexual males and females in order to provide additional evidence for the validity of unobtrusively measured VT as a measure of sexual preference. Both studies showed repeatable and consistent findings that subjects had the longest VT toward adult models of the preferred objects that decreased with the age of the model. VT for the non-preferred objects of all ages generally flattened out. The authors found general concordance between penile tumescence, attractiveness ratings, and VT. Quinsey

et al. state that "Viewing time offers advantages over penile plethysmography and rating methods" (p. 353).

Based upon past and recent research, it appears that VT can be used to determine sexual preferences or levels of sexual interest (Harris et al., 1996; Rosenzweig, 1942; Zamansky, 1956). Further, there is a positive correlation between VT and preferred sexual objects (Quinsey et al., 1996; Wright & Adams, 1994). However, there are some concerns regarding VT. For example, non-explicit stimuli may result in less variable and less distinct VT responses, and sexual offenders seem to show flatter, more restricted VT patterns (Quinsey et al., 1993). Notwithstanding these concerns, research supports the potential for VT assessment as a reasonable and meaningful alternative to PPG.

The Abel Assessment of Sexual Interest

The *Abel Assessment of Sexual Interest (AASI)* is a psychometric instrument that is designed to measure sexual attraction using 22 categories of possible sexual stimuli. During administration, participants view a total of 160 slides. Each slide contains a picture of an individual, each varying in age and gender. Some of the pictures portray various paraphilias, but all subjects are clothed. While viewing these slides, the subject is instructed to personally rate how sexually arousing each slide is using a 7-point Likert-type scale. Sustained attention to each slide is measured implicitly throughout this process (Fischer, 2000).

The *AASI* appears to be very promising; however, a number of questions have been raised regarding the validity and reliability of this assessment instrument. It claims to be able to measure attraction to a broad number of stimuli. The *AASI* avoids some ethical issues by not using images that are sexually explicit. It is not physically invasive,

and is less susceptible to faking because the examinees are not aware of the variable being measured.

To date, attempts to substantiate the reliability and validity of the *AASI* have produced discouraging results. Several researchers have examined the test-retest reliability of the *AASI*. Smith and Fischer (1999) found an average test-retest reliability of +.63. Kaufman, Rogers, and Daleiden (1998) found average coefficients ranging from +.55 to +.67. In both cases, the reliability did not fall in the desired range of $r > .80$ (Anastasi, 1988).

Research on the validity of the *AASI* has found mixed results. Smith and Annon (1998) compared the *AASI* to PPG and found no significant correlations between VT and admitted sexual behavior. Smith and Fischer (1999) examined the validity with adolescent participants and found similar results. In the Smith and Fischer study, the *AASI* was only able to discriminate offenders from non-offenders slightly better than what could be done by chance, with reported hit rates ranging from 52-59%. On the other hand, some researchers have found that the Abel Screen is able to identify offenders with accuracy comparable to PPG (Gray, 1999; Johnson & Listiak, 1999; Letourneau, 1999). Because these results are conflicting it is questionable as to whether the *AASI* should be used in the assessment of sexual offenders. Such results led Fischer and Smith (1999) to conclude that the *AASI* "is a promising instrument based on an interesting concept. However, the evidence of its reliability and validity for use with adults is weak as yet. Further refinement is necessary so that its use may become a reliable and valid means to promote appropriate treatment of sexual offenders and possibly also a means to protect potential victims" (pp. 203-204).

Affinity 2.0

Affinity 2.0 is a computer program that assesses sexual interest through a combination of overt self-report and covert measurement of sustained visual attention to stimuli (Glasgow, 2003). The *Affinity* program was initially developed to assess the sexual interest of men with mild mental retardation. However, the manual for the program, version 2.0, states that the program can be used for individual assessment with all adult male offenders and can be used for research purposes with adult male non-offenders, juvenile male offenders, and female offenders (Glasgow, 2003).

Assessment procedure. The *Affinity* program consists of two main tasks—a ranking procedure and a rating procedure (Glasgow, 2003). The ranking task consists of the subject being shown eight simple sketches of people representing both male and female adults, juveniles, pre-juveniles, and small children. Because the sketches are simple, clothing type, size, and ratio of body parts are the only discriminating factors. Subjects are directed to rank these images most sexually attractive to most sexually unattractive. The program is capable of displaying self-reported interest and actual viewing time, making discrepancies between the two immediately apparent visually.

During the rating task portion of the assessment, the participant is presented with images from each one of the eight stimulus categories (Glasgow, 2003). Each category depicts individuals of a certain gender and age group (adult, juvenile, prejuvenile and small child). The examiner can choose to have these 56 images presented randomly or in a predetermined sequence. Typically, a random administration is used. The subject is prompted to rate, on a Likert-type scale, how sexually attractive or unattractive they find the individual depicted in each slide. Possible ratings range from -7 — 0 — +7.

The computer records the subject's ratings while two types of VT are covertly measured, On Task Latency (OTL) and Post Task Latency (PTL). OTL is a measurement of viewing time, beginning with presentation of the image and ending when the participant makes a rating. PTL is a measurement of the time that elapses between the subject making a rating and then clicking "next image" (Glasgow, 2003).

Data interpretation. The results of *Affinity* can be presented as raw data—milliseconds per slide—or mean rank data. Mean ranks are the primary method for reporting results and are obtained by assigning each image a rank corresponding to its raw score. For OTL, the image that was viewed longest would receive a ranking of 56, while the image viewed for the shortest duration would receive a ranking of 1. Once all images have been ranked, a mean rank is computed for each category. Then the results are presented in a line graph (Glasgow, 2003).

The manual for *Affinity* lists two reasons for using mean rank, or ipsative data. First, it makes it possible to chart all data on a single graph. Second, it helps minimize the effect of outliers on mean scores (Glasgow, 2003). On the other hand, the manual acknowledges that

> A disadvantage of using mean ranks is that all correspondence with actual values in the raw data is lost, as is most sense of the distribution of scores. There is no "calibration" of latencies, so the absolute difference between latencies regarding two categories is only expressed relative to all the other scores, not to absolute duration. It is wise therefore to refer back and forth between raw data and mean rank charts in order to properly analyze the results. (Glasgow, 2003, p. 53)

According to Glasgow (2003), the preferred method of data interpretation is visual analysis, with attention to the real world significance of patterns to the data. Glasgow, Croxen, & Osborne (2003) give several examples of data interpretation that generally involve the visual analysis of results in linear chart form. The individual's history of sexual offense and sexual behavior is considered, as well as his self-reported attraction to certain stimulus categories, and how this information fits in with their pattern of viewing time. Particular emphasis is given to the mean rank viewing time information, because this information is considered to provide insight that is unaffected by dissimulation. If there were an inconsistency between an individual's viewing time pattern and their history or self-reported sexual interests, it would be justifiable to further examine the individual's self-reports, and consider the possibility of deceitfulness.

Reliability and validity. Research examining the reliability and validity of *Affinity* is limited. Glasgow et al. (2003) conducted a pilot study of *Affinity*, utilizing a single case methodology. They concluded that *Affinity* was able to produce valuable information as part of an assessment of sexual interest. However, because single case methodology does not lend itself to generalization, little evidence of reliability and validity was gained.

A second study by Glasgow and Croxen (2003) provided more information about validity and reliability. The sample for this study consisted of 31 pedophile offenders and 31 non-offenders. Internal consistency for each of the stimulus categories was analyzed, and the resulting Cronbach Alphas ranged between 0.76 and 0.93. Validity evidence was then examined in two ways. First, the researchers computed the correlations between rated sexual interest and viewing time. This analysis resulted in a median correlation of 0.57 for the non-offenders and 0.35 for the offenders. The authors accounted for this

difference by pointing out that there is a greater likelihood of "denied pedophile interest in the offender sample" (Glasgow & Croxen, 2003, p. 6). Second, they examined the discriminate validity of *Affinity* and found that the test was able to correctly identify pedophiles with 96% accuracy, while at the same time falsely identifying 23% of the non-offenders as offenders. In the authors' opinion, these results were sufficient to conclude that *Affinity* can be of practical use. These results are encouraging for a pilot study, but the reliability and validity of *Affinity* clearly must be examined further.

A Brief Introduction to Ipsative Scores and Their Weaknesses

Norm-referenced tests and assessments allow the practitioner to compare the participant's test results to those of others who have taken the same assessment. This is done in order to objectify an otherwise subjective domain and to aid the practitioner in interpreting data, thus assisting in the decision making process of determining deviance. Raw scores and standard scores are commonly used in norm-referenced psychological assessment. Therefore, practitioners are usually familiar with the basic characteristics and uses of this type of data. However, the mean rank scores generated by *Affinity* are uncommon, in that they describe a person's performance in comparison to himself instead of a comparison to others. Ipsative scores are used to make intra-individual comparisons. *Affinity* scores are considered to be ipsative. A technical definition of ipsatives is "any score matrix, which has the property that the sum of the scores over the attributes for each of the entities is constant" (Clemans, 1956). This means that ipsative scores always sum to a constant. If one score is higher on an ipsative scale, then one or more other scores must be lower. In other words, "each score for an individual is dependent on his scores or other variables" (Clemans, 1956).

One practical value of ipsative scores is that they allow a quick comparison of several attributes. However, a number of drawbacks exist. First, no inter-individual comparisons are possible and yet, ipsative scores are often misinterpreted as norm-referenced scores (Fischer & Smith, 1999; Johnson, Wood & Blinkhorn, 1988). Any attempts at inter-individual comparisons would be like concluding that because my right ear is larger than my left, then my right ear must also be larger than your left ear.

Second, 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.

Third, and most germane to the present study, conversion of scores from raw values to ipsative values often distorts the pattern of scores within the profile. Large differences between scores can be minimized and minor differences can be exaggerated in the conversion process. For example, if an individual has minimal variance in the raw score profile, forcing the scores to sum to a constant can force the pattern to show a degree of definition that is unsupported in the raw data. Figure 2 represents the actual raw score profile and the ipsative conversion of a subject in Harmon's (2007) sample of females. Her true underlying behavior manifested very little variance but the ipsative conversion made it appear as though there was a well-defined pattern of viewing time. If

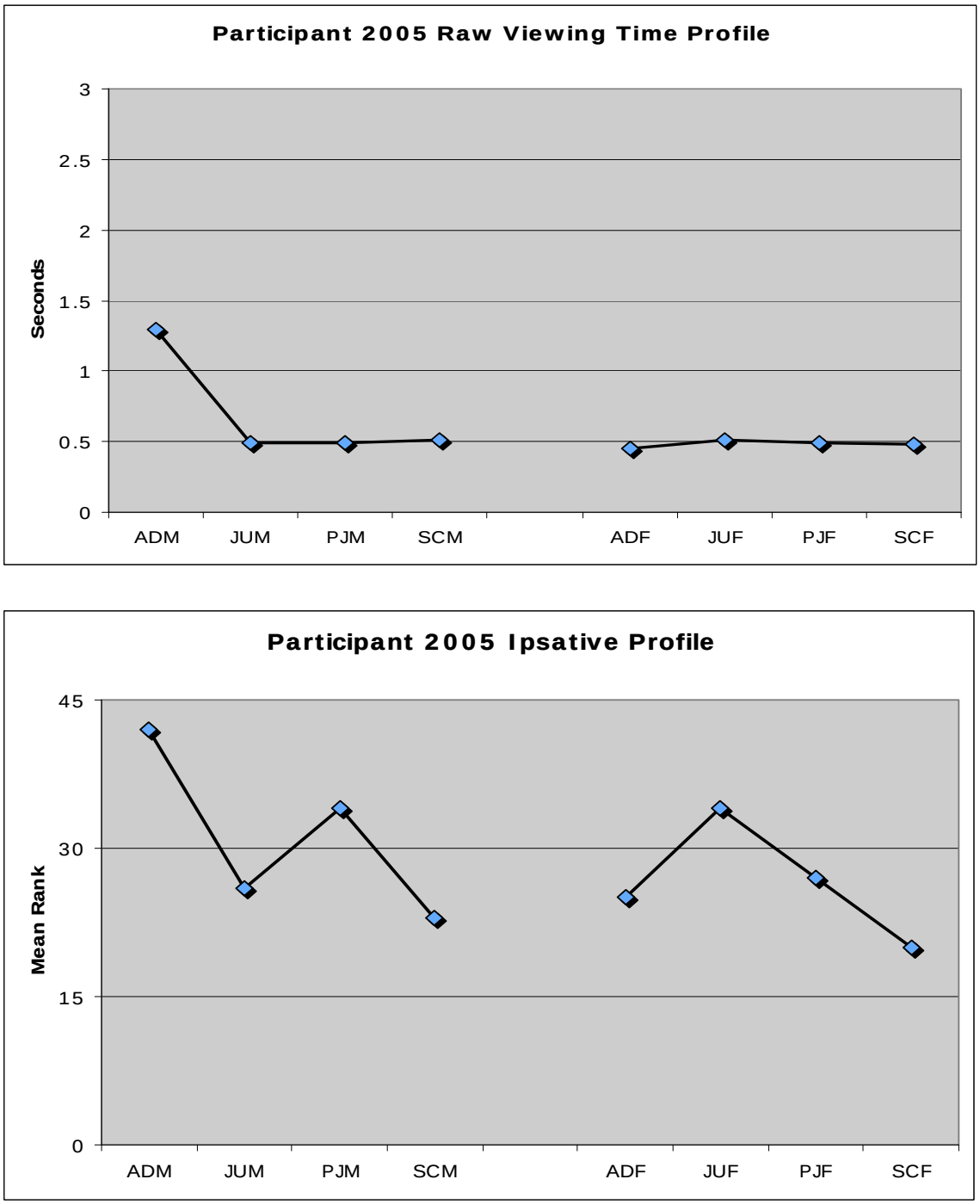


Figure 2. Raw score and ipsative conversion.

a practitioner were to rely solely on the mean rank scores for data interpretation on a case like this, then it is likely that they would reach false conclusions about the subject's sexual preferences.

Brown (2005) developed a method to measure distortion objectively in the conversion of *Affinity* raw scores among Harmon's (2007) sample of 146 non-offending, exclusively heterosexual females. He obtained distortion scores by ranking the ipsative scores on a scale from 1 to 8, with 1 being the category with the highest ipsative score and 8 being the category with the lowest score. Raw viewing times were also ranked in the same way, with the category with the longest VT being ranked 1, and the category with the shortest VT being ranked 8. After the scores had been ranked, the total number of position movements between raw VT and ipsative scores was totaled. Each value was converted to an absolute value of movement. One point was assigned for each rank position moved. In the case where there was a tie between two mean ranks, both ranks received half of the point total. This caused summed distortion scores to have odd values (1,3,5 . . .). The points for all categories were summed to obtain a total distortion score for that participant. An example of this procedure is shown in Table 1.

Using this method, Brown (2005) determined that 86% of all profiles incurred some degree of distortion. He formulated and tested several hypotheses as to variables that might be related to distortion. He experimented with composites of the variance and skew in the underlying raw score distributions. Brown's best predictor variables are shown in Table 2.

Table 1

Example of Method Used to Calculate Distortion Score by Comparing Rank Orders

	ADF	ADM	JUF	JUM	PJF	PJM	SCF	SCM	∑
Raw VT Rank	1	2	4	3	5	6	7	8	_
Ipsative Rank	2	4	1	3	5	6	7	8	_
Positions Moved	-1	-2	+3	0	0	0	0	0	_
Absolute Value of Positions Moved	1	2	3						6

Note:
ADF = Adult Female, ADM = Adult Male, JUF = Juvenile Female,
JUM = Juvenile Male, PJF = Prejuvenile Female, PJM = Prejuvenile Male,
SCF = Small Child Female, SCM = Small Child Male.

Table 2

Explanation of Variables Hypothesized to be Related to Distortion

Variable Label	Description
STM	The standard deviation of mean raw viewing times for each participant. This variable was calculated as a percent of the mean raw VT to correct for differences in mean raw VT from participant to participant. STM = (SDrawVT/MrawVT)*100
SDSKEW	The standard deviation of raw viewing time skew scores for a participant.
AVGSD	The average raw viewing time standard deviation across all eight stimulus categories. This variable was calculated as a percent of the mean raw VT to correct for differences in mean raw VT from participant to participant. AVGSD = [∑(SDADF, SDADM, . . .)/8]/MrawVT*100
SUMSKEW	The sum of raw viewing time skew scores for all eight stimulus categories. SUMSKEW = ∑(SKEWADF,SKEWADM, . . .)

Brown (2005) then used those predictors to account for the variance in his distortion values. The relationships were examined using Pearson Product Moment Correlations. The relationship between STM and distortion was not significant (r = -.162, p > .05). The predictor AVGSD was significantly related to distortion (r = .405, p < .01). The variable SUMSKEW was significantly related to distortion (r = .321, p < .01). The variable SDSKEW was also significantly related to distortion (r = .265, p < .01). While three of Brown's individual variables were significant predictors of distortion, no one predictor was outstanding.

Attempting to enhance the prediction of distortion, Brown calculated several interaction effects using different combinations of the variables. The strongest correlation was found when all four variables were used to calculate an interaction effect. This interaction effect (INT) was calculated using the following formula:

$$INT = (AVGSD*SUMSKEW*SDSKEW)/STM$$

INT was found to have a moderately strong correlation with distortion (r = .530, p < .01) and accounted for 28.1 % of the total variance in distortion scores. Brown (2005) developed an algorithm for identifying significant distortion using the interaction effect. He tested several cut-off scores and the most effective cut-off correctly identified 42.9% of the significantly distorted profiles.

Statement of Problem

When raw scores from *Affinity* are converted to ipsative scores, they are changed from ratio to ordinal data. In doing so, the data's relation to actual viewing times is lost. Theoretically, certain distortions can occur as a result of this conversion—possibly

causing insignificant differences in viewing time to appear meaningful, or even minimizing meaningful differences.

Attempting to further examine the reliability of the *Affinity* assessment—in order to demonstrate the need for interpretation and decision making based on norm-referenced data—researchers at Brigham Young University recently gathered data on samples of non-offending, exclusively heterosexual males and females. While doing so, they noticed in that the conversion from raw viewing times to ipsative scores resulted in distortion of the patterns. The ipsative score profile results in a pattern of sexual attraction differing from that of the raw viewing time profile. Being able to warn clinicians of the potential invalidity of a profile would decrease the probability of drawing erroneous conclusions regarding an examinee, and possibly indicate cases where the examinee was trying to obscure their true sexual preferences.

Statement of Purpose

The purpose of this study was to test Brown's (2005) model on the responses of 99 non-pedophilic, exclusively heterosexual males to *Affinity 2.0*. All profiles were objectively measured and examined to identify distortion using Brown's analytical technique. The same variables (STM, SDSKEW, AVGSD, SUMSKEW) Brown hypothesized would predict distortion and their interaction effects were applied to the male sample.

METHODS

Participants

The present study utilized and analyzed data collected by Crosby (2008) as a part of his doctoral dissertation. His study examined the reliability of *Affinity 2.0* scores with non-offending, heterosexual males. Participants were 99 males with a minimum age of 18. All participants were undergraduate students at Brigham Young University and were recruited using short presentations given by researchers to undergraduate psychology classes. Potential participants were informed in these presentations that the purpose of the study was to test a new device that measured sexual interest. They were also informed that participation would involve viewing several pictures of fully clothed models, rating those images on their attractiveness or unattractiveness, and then completing a short questionnaire. Participants were compensated with movie tickets and/or extra credit.

Only participants with no history of pedophilia, who reported exclusive heterosexual interest, were included in the study. Participants were screened for these variables through the informed consent procedure and a self-report questionnaire that was administered at the end of the evaluation.

Measures

All participants included in the experimental group completed *Affinity 2.0* at test and retest, although only the time one data was analyzed. *Affinity* is a computer program that assesses sexual interest through a combination of overt self-report and covert measurement of sustained visual attention to various stimuli (Glasgow, 2003). The *Affinity* program was initially developed to assess the sexual interest of men with mild mental retardation. However, the manual for version 2.0 of *Affinity* states that the

program can be used for individual assessment with all adult male offenders, and can be used for research and evaluation with adult male non-offenders, juvenile male offenders, and female offenders (Glasgow, 2003).

The *Affinity* program consists of two main tasks—a ranking procedure and a rating procedure (Glasgow, 2003). The ranking task allows the participant to self-report the stimulus groups they find sexually attractive. During the ranking task the participant is exposed to images depicting individuals of varying ages and genders. The participant is asked to rate how sexually attractive or unattractive he finds the individual depicted in the picture on a Likert-type scale. The participant's ratings are recorded by the computer, while viewing time is measured (Glasgow, 2003).

After taking *Affinity 2.0* for the second time, participants were given the Demographics, Social Desirability and Sexual Interest Questionnaire (DDSQ). The DDSQ is composed of two sections. In the first section, the participants reported demographic variables such as age, ethnicity, year in school, and marital status. The second section of the DDSQ contained a question about sexual orientation. The question was an adaptation of the Kinsey Heterosexual-Homosexual Scale (Kinsey, Pomeroy, & Martin, 1998). This question was used as an exclusionary variable.

Procedures

Assessing sexual interest can be an intrusive and uncomfortable experience for many individuals. Some individuals may be hesitant to respond to questions honestly unless a comfortable and confidential environment is provided. Brigham Young University is a private religious school that requires students to comply with a strict code of conduct that prohibits sexual activity outside of marriage. This includes: homosexual

activity, viewing pornography, masturbation, pedophilia, and extra-marital sexual relations. Violations of this conduct code can result in university sanctions (Brigham Young University, n.d.). Consequently, participants with homosexual or pedophilic tendencies may be extremely hesitant to respond honestly on research instruments.

In order to address this issue, all participants were required to read and sign an informed consent document. Participants were provided with a comfortable, confidential assessment environment. The informed consent document explained the purpose of the study and expectations for the participants. It also explicitly discussed the methods for protecting the confidentiality of the identity of all potential participants. This included assigning numbers to each participant, deleting names, only keeping one master list of the names, and ensuring the participants that no names would be used in the study or reported to the university. The master list was kept in a secure, locked file.

The informed consent procedure was also used to screen for any history of pedophilia. A negative history of pedophilia and an age of 18 or older were the initial inclusion criteria. Participants were asked to confirm that they met those criteria before engaging in the assessment.

After reading and signing the informed consent document, participants were taken to a private room, containing a computer loaded with the *Affinity 2.0* program. With each participant, the researcher instructed the subject on how to begin the program, and then exited the room. To make sure that they were not disturbed, the researcher informed each participant that they would wait outside the room until the participant completed the assessment. After completing the assessment, the researcher answered any participant

questions and scheduled a time to retest. At retest, the same procedure was followed with the exception that the DDSQ was administered following *Affinity 2.0*.

Data Analysis

Brown's (2005) analytical methods were applied to the time-one profiles of Crosby's (2008) sample of male responses to *Affinity 2.0*. First, distortion scores were created for each participant. The raw viewing time and ipsative score profiles were graphed for each participant using Microsoft Excel. These graphs are contained in the Appendix. Then Brown's predictors were calculated. Next, the predictors were correlated individually with the distortion scores. Finally, Brown's interaction term was calculated and correlated with the distortion scores.

RESULTS

Brown's (2005) method described in the data analysis section was utilized in order to calculate distortion scores for each participant. Frequencies of distortion scores are outlined in Table 3. When this method was applied to the sample, 12 participants had a distortion score of 0 or 1. For the remaining 87 participants, distortion scores ranged from 2 to 22.

Next, Pearson Product Moment Correlations were calculated for the variables in Table 2. The correlation between distortion and STM was not significant ($r = .085$, $p > .05$). The correlation between distortion and SDSKEW was not significant ($r = .174$, $p > .05$). The correlation between distortion and AVGSD was not significant ($r = .154$, $p > .05$). The correlation between distortion and SUMSKEW was not significant ($r = .139$, $p > .05$). With a female sample, Brown (2005) found three out of his four variables to be statistically significant. In the present study, the three variables AVGSD, SDSKEW, and SUMSKEW were also the most strongly correlated with distortion. Out of the four, however, none were found to be statistically significant.

Brown's (2005) interaction term, INT, was correlated with distortion and resulted in a weak but significant zero-order correlation of .248 ($p < .05$) which accounted for only 6% of the variability ($r^2 = .06$). The relationship between the variable INT and distortion is graphically represented in Figure 3. Because of the very poor prediction of distortion by the individual predictors and the interaction term, no attempts were made to create an algorithm to screen significantly distorted profiles.

Table 3

Frequency of Distortion Scores

Distortion Score	Frequency	Percent
0	11	11
1	1	1
2	15	15
3	3	3
4	14	14
5	4	4
6	20	20
7	1	1
8	13	13
10	6	6
12	6	6
14	2	2
16	1	1
18	1	1
22	1	1

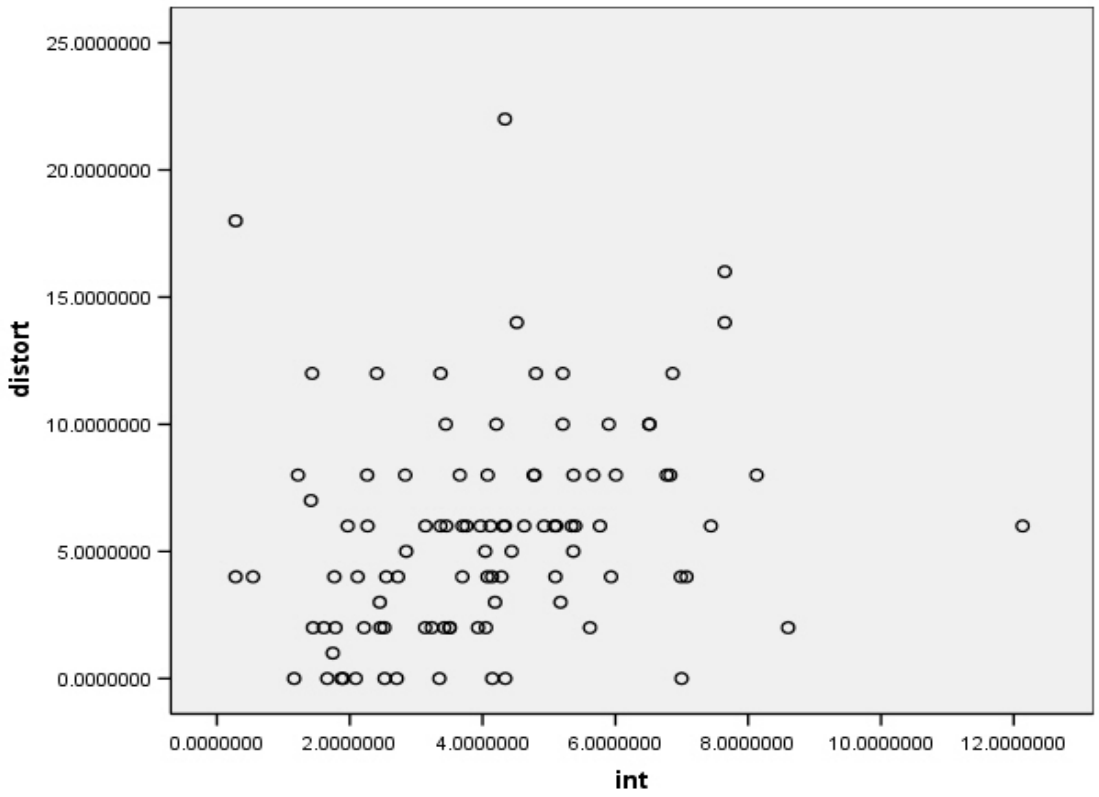


Figure 3. Scatterplot of distortion scores and interaction effect values for entire sample.

DISCUSSION

The primary purpose of the present study was to test Brown's (2005) model among the time-one sample (Crosby, 2008) of 99 non-pedophilic, exclusively heterosexual males and their responses to *Affinity 2.0*. All profiles were objectively measured and examined to identify distortion using Brown's (2005) analytical technique. The same variables that Brown hypothesized would predict distortion and their interaction effects were applied to the present study. The variables related to distortion were tested, and none were found to have positive correlations with distortion. Even when interaction effects were calculated, only a weak relationship was found.

One possible reason that the males' responses did not result in a robust prediction of distortion may be related to the differences in male and female sexuality and the way in which the predictors, particularly INT, was calculated. In general, it has been shown that objective assessment of female sexuality is quite unreliable (Heiman, 1977 & 1980; Hoon, 1984; Wincze, Hoon & Hoon, 1978). It may be that females' underlying distributions tend to be much more variable and skewed than males' distributions. In fact, the range of Brown's (2005) INT scores was from 1 to 58. The range of INT scores in this male sample was 1 to 12. Given that INT is a combination of skew and standard deviation estimates, smaller skews and standard deviations (associated with less variability) would result in a smaller range of INT values. If so, a restricted range of INT values would tend to lead to lower correlations between INT and distortion. In some ways, this seems paradoxical in that the greater variability in females' distributions leads to greater predictability of the distortions in their scores. Conversely, the reduced

variability among the males' distributions leads to decreased prediction of their distortion scores.

Although the correlation coefficients were significant with the female sample, Brown (2005) emphasized that those relationships should not be overstated. The variables only appeared to account for a small portion of the variability in the distortion scores, and concluded that there must be other factors, yet undiscovered, that were more significant contributors to distortion. The variables measured with the males accounted for even less of the variability than in the female sample and were unable to predict distortion among a male sample.

In Brown (2005), only 21 of the 146 participants, or 14.4%, had zero distortion between their raw viewing time profiles and their ipsative score profiles. The remaining 125 participants (85.6%) had at least some ipsative score distortion. In the present study, only 11 of the 99 participants, or 11%, had zero distortion between their raw viewing time profiles and their ipsative score profiles. The remaining 88 participants (88.9%) had at least some ipsative score distortion. For the purposes of his study, Brown defined significant distortion as a distortion score greater than 10. He described how the cut-off was rather arbitrary, and how the argument could be made that all distortion is meaningful to the practitioner. For example, if a participant spent the most time viewing slides in the adult female category, followed by the adult male category, but these category rankings were inverted on the ipsative score profile; this would only be a distortion score of 2 (assuming all other category placements remained consistent). Although this is a low distortion score, even this minor distortion could lead to contrary

conclusions about the client's preferred sexual stimuli (i.e., he is more sexually attracted to adult males than to adult females).

The use of ipsative scores is questionable because they may lead to erroneous conclusions inconsistent with the raw viewing time profile. As mentioned previously, over 88% of the profiles in the present study were distorted to some degree, which means that the ipsative profiles inaccurately represent sustained visual attention in the majority of cases. Research supports sustained visual attention as an indicator of sexual interest, but ipsative scores do not accurately represent visual attention, thus threatening the validity of the instrument.

It may be argued that this is over-exaggerating the flaws, but the evidence from both Brown (2005) and the present study demonstrates that distortion is the rule rather than the exception with *Affinity 2.0*. This should not be interpreted to mean that the underlying theory is unsound or that the instrument itself should be discontinued. However, since a reliable formula for identifying distortion has not been found and distortion seems too prevalent, alternative methods (i.e., norm-referenced) of scoring should be examined. The present study and Brown's seem to underscore the value of such exploration. Since all of the foundational research used raw VT rather than ipsative scores, developing a scoring method less susceptible to distortion, and more closely tied to raw VT, would be ideal. If alternate methods for scoring *Affinity* can be found, then it would be unnecessary to "flag" distorted profiles.

In an effort to address a limitation identified in Brown (2005), the purpose of the present study was to investigate whether a similar pattern of distortion would occur in a male sample as it did with females. The present study found that there was slightly more

distortion among the male population, but that the variables hypothesized to cause the distortion were less likely to be identified as predictors than was found with the female population.

Based on the present study and Brown (2005) there are several future studies that could be conducted. These ideas could be tested among samples other than exclusively heterosexual people. It would be important to know what the distortion rates are among other groups. Alternative raw score markers that might account for both male and female score distortion could also be investigated and tested in order to predict where distortion would occur. Also, a comparison of male and female raw score distributions may inform the field about the differences between male and female viewing time behavior.

REFERENCES

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

Anastasi, A. (1988). *Psychology testing* (6th ed.). New York: Macmillan.

Brigham Young University. (n.d.). Church educational system honor code. Retrieved July 29, 2006 from http://campuslife.byu.edu/HONORCODE/chaste.htm

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

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

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

Cloyd, L. (2006, September). *The correlation between Monarch PPG & Affinity.* Paper presented by 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.

Crosby, D. (2008). *Non-pedophilic heterosexual male response to the Affinity 2.0.* Unpublished doctoral dissertation, Brigham Young University, Provo, UT.

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

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.

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

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

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

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

Harmon, K. (2007). *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.

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

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

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

Johnson, S. A., & Listiak, A. (1999, September). Comparing outcomes of plethysmographic assessment with the Abel assessment in a prison based sex offender sample. In S. Johnson (Chair), *Current research comparing plethysmography with the Abel Assessment for Sexual Interest.* Symposium conducted at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, FL.

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

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

Laws, D. R., & Gress, C. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9* (2), 183-196.

Letourneau, E. J. (1999, September). A comparison of the penile plethysmograph with the Abel Assessment for Sexual Interest on incarcerated military sex offenders. In S. Johnson (Chair), *Current research comparing plethysmography with the Abel*

Assessment for Sexual Interest. Symposium conducted at the annual meeting of the Association for the Treatment of Sexual Abusers, Orlando, Fl.

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.

Quinsey, V. L., Rice, M. E., Harris, G. T., & Reid, K. S. (1993). The phylogenetic and ontogenetic development of sexual age preference 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.

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

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

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

Smith, B., & Annon, J. (1998, October). *Abel assessment and penile plethysmography: A comparison of instrument sensitivity.* Poster session presented at the annual meeting of the Association for the Treatment of Sexual Abusers, Vancouver, British Columbia.

Smith, G., & Fischer, L. (1999). Assessment of juvenile sex offenders: Reliability and validity of the Abel Assessment for interest in paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 1* (3), 207-216.

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., & Pruesse, M. (1972). The semantic meaning of pornographic stimuli for college males. *Canadian Journal of Behavioral Science, 4*, 204-209.

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

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

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

APPENDIX