Dissertations, Theses, and Capstone Projects   CUNY Graduate Center

9-2015

A Cognitive-Based Indicator of Deviant Sexual Arousal: Concurrent Validation of the Emotional Stroop

Ashley H. Spada
Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

A COGNITIVE-BASED INDICATOR OF DEVIANT SEXUAL AROUSAL:

CONCURRENT VALIDATION OF THE EMOTIONAL STROOP

By

Ashley H. Spada

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015

This manuscript has been read and accepted for the
Graduate Faculty in Psychology in satisfaction of the
Dissertation requirement for the degree of Doctor of Philosophy.

_____   _____
Date   Elizabeth L. Jeglic, Ph.D.
                                               Chair of Examining Committee

_____   _____
Date   Joshua Brumberg, Ph.D.
                                               Executive Officer

Elizabeth L. Jeglic, Ph.D.

Cynthia Calkins, Ph.D.

Peggilee Wupperman, Ph.D.

Charles Amrhein, Psy.D.

Pietro Lofu, Psy.D.

_____
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

A COGNITIVE-BASED INDICATOR OF DEVIANT SEXUAL AROUSAL:

CONCURRENT VALIDATION OF THE EMOTIONAL STROOP

by

Ashley H. Spada

Adviser: Professor Elizabeth L. Jeglic, Ph.D.

One of the strongest dynamic predictors of sexual recidivism among sex offenders is deviant sexual arousal (DSA; Hanson & Bussiére, 1998). Phallometric testing, the most commonly used method of assessing DSA, has elicited numerous methodological, ethical, and financial criticisms, while self-report measures are vulnerable to social desirability and lack of self-awareness. In an effort to overcome the limitations of previous measures of DSA, researchers have employed cognitive measures including a modified version of the Stroop task to measure DSA among sexual offenders (Price & Hanson, 2007; Smith & Waterman, 2004). These original studies used victim selection to assess the concurrent validity of the Stroop as a measure of DSA. However this assumes that victim selection is based upon DSA, which is not always the case. Therefore the research to date evaluating the Stroop as a measure of DSA has been inconclusive at best. Thus, the current validation study explored the ability of the Stroop task to assess DSA by comparing deviant Stroop latency to self-reported DSA among a non-offending sample of 570 females and 223 males. The results indicated a significant relationship between self-reported deviance and deviant Stroop performance, suggesting that the Stroop is a viable option for assessing DSA. The results have far reaching implications for sex offender risk assessment, as validation of a non-invasive, cognitive measure would allow clinicians to more quickly and ethically assess sexual offenders for risk of reoffending by evaluating their level of DSA.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my advisor, Elizabeth Jeglic, for her many years of mentorship, patience, dedication, and overwhelming support throughout my academic career. She has always provided the perfect balance of support, guidance, and gentle pushing to challenge myself to fully develop my professional potential.

I would also like to express my gratitude to the many fellow students and research assistants who contributed to this dissertation project and without whom this endeavor would not have been possible. Yeunjoo Chung, Julia Mesler, Alessandra Gallo, and Ashley Rupp, I sincerely thank you for your hard work and dedication. I would like to thank The Graduate Center of the City University of New York for supporting this project with their Doctoral Student Research Grant Program.

I would also like to thank my family, friends, and colleagues for their endless support as I worked toward achieving my professional goals.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: Introduction & Scope

Sexual offenses are among the crimes that are most publicized by the media and represent a grave concern to the public (Hanson & Morton-Bourgon, 2005; Levenson, Fortney, & Baker, 2010). According to the Bureau of Justice Statistics, there were 346,830 instances of reported sexual assault in the United States in 2012, the most recent year for which national crime victimization data is available (Truman, Langton, & Planty, 2013). Sensationalized media coverage of sexual crimes has contributed to a widespread perception that sexual offenders are violent, predatory, and highly likely to reoffend (Finkelhor, 2009).

In actuality, sexual offenders are comprised of a diverse group of individuals who vary in their re-offense risk (Finkelhor, 2009). In order to address an increasing public concern about sexual recidivism, legislators have created several new policies over the past twenty years in an effort to prevent instances of future reoffending (Levenson et al., 2010; Blasko, Jeglic, & Mercado, 2011; Zgoba & Simon, 2005). Despite the growing body of legislation geared toward sexual offenders, it should be noted that overall rates of sexual recidivism are relatively low compared to recidivism rates for other types of offenders, and often significantly lower than anticipated by the public (Hanson & Bussiére, 1998; Hanson & Morton-Bourgon, 2005). Research has consistently demonstrated that a *minority* of sexual offenders reoffend, with detected rates of sexual recidivism ranging from 5% to 14% over a follow-up period of approximately three to five years after prison release (Hanson & Bussiére, 1998; Hanson & Morton-Bourgon, 2009). Further, the majority of sexual offenses (95%) are committed by first-time offenders, not by those who have been previously incarcerated for a sex crime (Sandler, Freeman, & Socia, 2008).

However, rates of recidivism are not uniform across sex offenders, as sex offenders are a heterogeneous and complex group that includes juvenile and adult offenders, those with and

without severe psychopathology, and those who pose varying degrees of risk for recidivism based on various static and dynamic characteristics (Finkelhor, 2009). Some research has suggested that sexual offenders with psychopathy are at higher risk to recidivate (Serin, Mailloux, & Malcom, 2001). Other studies have suggested that the sex offenders at greatest risk for recidivating include rapists as compared to child molesters, sex offenders with extrafamilial and stranger victims as opposed to incest offenders, and those who molest young males (Hanson & Bussiére, 1998; Harris, Rice, Quinsey, Lalumiére, Boer, & Lang, 2003; Levenson et al., 2010).

However, there are even limitations to the way in which we categorize sex offenders, as these categories are typically based on the age and relationship to the victim (e.g., incest offender, child molester, rapist), thus implying that sex offenders are consistent in their victim choice (Kleban, Chesin, Jeglic, & Mercado, 2012). However, the research has been somewhat mixed with regard to the stability of victim choice, and this is extremely problematic in terms of predicting sexual recidivism.

Kleban and colleagues (2012) found that while 46% of sex offenders with multiple victims involved in their index offense offended exclusively against family members, 26% exclusively against acquaintances, and 11% exclusively against strangers, 14% offended against both family members *and* acquaintances or against both family members *and* strangers. While 52% maintained a consistent victim type across their sex offense histories, 48% offended against victims of various degrees of relationship to the offender (e.g., strangers, acquaintances, family). Further, 17% had both male and female victims, and 23% offended against victims of various age ranges (Kleban et al., 2012). These findings suggest that perhaps sex offenders may crossover in terms of victim age and relationship but remain stable in terms of victim gender.

Given the complexity of sexual offending and recidivism, it is important to consider a multitude of factors and characteristics when assessing risk for sexual recidivism.

An early approach to risk assessment was to consider sexual offenses as similar to other types of criminal behavior and to therefore investigate the factors that predict general recidivism. The fact that many sex offenders engage in nonsexual criminal offenses, in addition to sexual offending, makes this approach to assessment intuitive and face valid (Hanson & Bussiére, 1998). The literature on general, nonsexual criminal behavior has identified a set of risk factors predictive of general recidivism. These factors include young age, unstable employment, alcohol and substance abuse, procriminal attitudes, and relationships with other criminals and are similar to the characteristics that comprise antisocial personality disorder according to the Diagnostic and Statistical Manual of Mental Disorders (*DSM-5*; APA, 2013; Gendreau, Little, & Goggin, 1996). While sex offenders are more likely to recidivate with a nonsexual crime than with a new sex crime, nonsexual criminals *rarely* recidivate with sex offenses. This suggests sexual offending may be fundamentally different from other types of criminal behavior (Hanson & Bussiére, 1998; Hanson, Scott, & Steffy, 1995). In fact, sex offenders known to be persistent and high risk may often score in the low risk range on these general recidivism measures. Therefore, it is not surprising that measures created to assess criminal recidivism in general have not been effective in predicting sexual recidivism.

Indeed, in their meta-analysis of 87 sexual recidivism studies, Hanson and Bussiére (1998) found that sexual criminal history was weakly related to general, nonsexual recidivism. While some factors that were predictive of general nonsexual recidivism also predicted sexual recidivism, including pro-criminal attitudes and failure to complete treatment, the *strongest* predictors of sexual recidivism were also *unique* to sexual reoffending. The strongest predictors

of sexual recidivism were deviant sexual interests, prior sexual offenses, and deviant victim choices such as boys, children, and strangers. This study provides support for the notion that general recidivism risk factors are not sufficient to predict sexual recidivism (Gendreau, Little, & Goggin, 1995; Hanson & Bussiére, 1998; Hanson, Scott, & Steffy, 1995).

Determinations of recidivism risk greatly influence numerous important decisions related to sentencing recommendations and proceedings, recommendations for civil commitment, and mandated treatment. Since sex offenders pose varying degrees of risk for reoffending, it is therefore crucial that risk assessment procedures accurately identify those at highest risk. This would allow policies, treatment, and limited resources to be targeted for the highest risk individuals, those for whom the interventions are most appropriate and cost-effective (Blasko, Jeglic, & Mercado, 2011; Finkelhor, 2009; Storey, Harris, 2014; Watt, Jackson, & Hart, 2012).

Given the poor predictive validity of general recidivism measures to predict sexual recidivism together with evidence suggesting that sexual offending might be different from other types of criminal behavior, researchers moved to investigate risk factors that were *specific* to criminal sexual behavior. Best practices in sexual recidivism risk assessment therefore currently includes a three-level assessment of factors that have been empirically linked to sexual recidivism including static factors, stable dynamic factors that can be targeted in treatment, and the more acute dynamic factors (Canales, Olver, & Wong, 2009; Hanson & Harris, 2001).

CHAPTER 2: Literature Review

Recidivism Risk Assessment

Static, historical factors are factors inherent to the individual and their offense history, and they are relatively fixed. Static factors such as offender age, relationship status (never having been in a long term relationship for two years), number of previous sexual convictions, relationship to victim(s), and number of male victims have been shown to be associated with sexual recidivism (Hanson, 2000; Hanson, Steffy, & Gauthier, 1993; Hanson & Morton-Bourgon, 2009). These factors have been investigated using various actuarial measures (e.g., Rapid Risk Assessment for Sex Offense Recidivism [RRASOR; Hanson, 1998], Structured Anchored Clinical Judgment [SACJ-Min; Hanson & Thornton, 2000], Static-99 [Hanson & Thornton, 1999], Minnesota Sex Offender Screening Tool-Revised [MnSOST-R; Epperson, Kaul, Huot, Goldman, & Alexander, 2003]).

The Static-99 (Hanson & Thornton, 1999) is the most widely used and accepted actuarial measure of static factors related to sexual recidivism risk (Storey et al., 2012). The Static-99 is a ten-item measure coded from archival data from an offenders' file, and the resulting scores categorize offenders into one of four risk groups: low, moderate, moderate-high, and high risk (Storey et al., 2012). Research has consistently demonstrated that Static-99 factors such as offender age, marital status, number of sexual offense convictions, relationship to victim(s), and number of male victims are highly associated with sexual reoffending (Hanson & Morton-Bourgon, 2009). On average, the Static-99 yields moderate to strong accuracy in predicting sexual recidivism (Hanson & Morton-Bourgon, 2009; McGrath, Cumming, & Burchard, 2003).

However, some more recent studies have revealed lower estimates of reliability and predictive validity when the Static-99 has been examined in more applied settings (Boccaccini,

Murrie, Caperton, & Hawes, 2009; Boccaccini et al., 2012; Storey et al., 2012). Further, a limitation inherent to the use of static factors to determine recidivism risk is that they are, by definition, static and unchanging over time, which renders them useless as indicators of change. Thus, despite the fact that treatment or other dynamic factors, such as alcohol consumption and/or disruption of stable living environment, may alter the actual risk for future reoffending, the Static-99 score will not increase or decrease to reflect the corresponding changes in risk (Hanson & Harris, 2001; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). Therefore, the identification of risk factors that fluctuate and may be potential treatment targets is a critical aspect of sexual recidivism prediction and prevention (Hanson & Harris, 2000).

Dynamic risk factors, also referred to as criminogenic needs, are characteristics that are currently present and amenable to change (Andres & Bonta, 1994). Identification of dynamic risk factors is most ideal, as they provide information about factors that can be addressed and potentially altered in treatment or by enhanced supervision (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012). Harris and Hanson (2010) define dynamic risk factors as skills deficits, personal predilections, and coping styles that can be meaningful intervention targets. Examples of dynamic risk factors include employment, drug abuse, psychological symptoms, and social supports/relationship status (Hanson & Harris, 2000). Hanson and Harris (2000) explain that dynamic risk factors can be further divided into two subtypes, stable and acute dynamic factors.

Acute dynamic risk factors are those that are time-limited and change rapidly. They may change over a period of weeks, days, or hours, and include factors such as alcohol or substance intoxication and/or acute distress (Hanson, Harris, Scott, & Helmus, 2007). A recommended

timeline for assessing acute factors is those that are present since the last time the offender was evaluated and no more than once per week (Harris, 2014).

Stable dynamic risk factors are those that may remain unchanged for extended periods of time (months or years) without intervention (i.e., psychiatric symptoms, sexual deviance, use of sex as coping) and are the subtype of dynamic risk factors most commonly associated with sexual recidivism (Hanson & Harris, 2000). Though they can be altered, they are thought to change slowly and only with much effort (Eher et al., 2012). Harris (2014) recommends a timeline of one year before and after the time of the evaluation when assessing an individual on the presence of stable dynamic risk factors.

Hanson and Harris (2007) created a three-tiered level of risk assessment that includes evaluation of static, stable, and acute factors using the Static-99 (Hanson & Thornton, 1999), Stable-2007, and Acute-2007 (Hanson, Harris, Scott, & Helmus, 2007). Based on previous research, they identified 13 stable dynamic factors predictive of sexual recidivism, including: social influences, intimacy deficits (capacity for relationship stability, emotional identification with children, hostility toward women, social rejection, and lack of concern for others), general self-regulation difficulties (impulsivity, poor cognitive problem solving, negative emotionality), sexual self-regulation difficulties (sexual preoccupation, sex as coping, deviant sexual interests), and lack of cooperation with supervision (Hanson & Harris, 2007).

Among their sample of 208 recidivistic and 201 non-recidivistic sexual offenders in Canadian correctional facilities, Hanson and Harris (2000) found that stable dynamic risk factors were more accurate in differentiating recidivists from non-recidivists than either the acute dynamic factors or the static factors. In their study, the dynamic factors found to be associated with recidivism included poor social support, attitudes of sexual entitlement, antisocial lifestyles,

and poor self-management. Further, dynamic risk factors remained strongly predictive of sexual recidivism even after controlling for several static factors (Hanson & Harris, 2000). The utility of dynamic risk factors in predicting sexual recidivism has since been replicated by other research groups (Eher et al., 2012; Olver, Wong, Nicholaichuk, & Gordon, 2007). Given that the research has consistently demonstrated dynamic factors, in general, to be empirically related to sexual reoffending, attempts have been made to identify the specific dynamic factors that *best* predict reoffending in an effort to detect those at highest risk.

The stable dynamic risk factor that has been most frequently empirically associated with sexual recidivism is deviant sexual arousal, defined as arousal to or interest in people, objects, or activities that are illegal, inappropriate, or highly unusual, usually assessed by behavioral history, self-report, criminal record, and the results of specialized testing (e.g., phallometric testing; Harris & Hanson, 2010). However, defining sexual deviance is somewhat controversial and warrants further discussion below.

Although sexual offenses of course qualify as deviant *behaviors*, it has been established that not all those who perpetrate sexual offenses have deviant sexual interests or arousal patterns. Sexual offenders who do exhibit a pattern of deviant sexual arousal appear to be more likely to reoffend than sex offenders who do not exhibit such patterns (Akerman & Beech, 2012; Hanson & Morton-Bourgon, 2005; Seto, 2001). In fact, deviant sexual interest was the single best predictor of sexual recidivism in the meta-analysis of 61 recidivism studies described above (Hanson & Bussiére, 1998).

Hanson and Harris (2000) found that among their sample of recidivists and non-recidivists, the recidivistic group was significantly more deviant in that they had more diverse victims with regard to age and gender, stranger victims, juvenile sexual offenses, and paraphilia

diagnoses than the non-recidivistic group. In fact, deviant sexual arousal, such as sexual interest in children, has been shown to be the strongest and most stable dynamic predictor of sexual recidivism (Hanson & Bussiére, 1998; Hanson & Morton-Bourgon, 2005). Sexual deviance ($r =$ .32) has exhibited stronger correlations to reoffending than all other variables linked to recidivism, including history of prior sexual offenses ($r = .19$), having stranger victims ($r = .15$), antisocial personality disorder ($r = .14$), prior nonsexual criminal offenses ($r = .13$), and having male victims ($r = .11$; Hanson & Bussiére, 1998). Meta-analyses have demonstrated this link between deviant sexual arousal and recidivism across various sexual offender subtypes (Barbaree & Marshall, 1998; Hanson Malcom, Andrews, & Quinsey, 1993; Quinsey, Rice, & Harris, 1995; Rice et al., 1990).

In addition to being highly correlated with sexual recidivism, sexual deviance is of particular importance as it can be directly targeted in treatment. The majority of current evidence-based sexual offender treatment programs employ a cognitive-behavioral framework and target sexual deviance, intimacy deficits, modification of cognitive distortions, social skills training, and relapse prevention strategies (Moster, Wnuk, & Jeglic, 2008; Olver, Nicholaichuk, Gu, & Wong, 2012). Despite an unpromising review of the early treatment outcome literature (Furby, Weinrott, & Blackshaw, 1989), more recent reviews and advances in therapeutic interventions for DSA and other criminogenic needs have demonstrated significantly more optimistic findings. In their meta-analysis of 43 sex offender treatment studies, Hanson and colleagues (2002) found that, among 9,000 offenders, treated sexual offenders were significantly less likely (12.3%) to reoffend than non-treated sexual offenders (16.8%), and that the interventions most strongly related to reductions in recidivism were cognitive-behavioral in nature. Other studies have demonstrated results consistent with Hanson and colleagues (2000),

demonstrating the efficacy of cognitive-behavioral, behavioral, and biomedical treatments in reducing sexual recidivism (Gallagher, Wilson, Hirschfield, Coggeshall, & MacKenzie, 1999; Hall, 1995). Even more recently, Olver and colleagues (2012) demonstrated that sex offenders who completed a comprehensive correctional treatment program targeting sexual deviance and several other sexual offense-related criminogenic needs recidivated at a rate significantly lower (10.7%) than the non-treated comparison sample (20.2%) over a follow-up period of 11.7 years.

Overall, treatment programs that address multiple dynamic targets including deviant sexual arousal and fantasy, cognitive distortions, and social competence appear to be most effective in treating sex offenders. Given the evidence that deviant sexual arousal is so closely related to sexual recidivism, taken together with preliminary evidence demonstrating the efficacy of targeting sexual deviance in treatment, sound assessment of deviant sexual arousal is critical so that treatment-related change may be accurately evaluated (Akerman, 2010).

Defining Deviant Sexual Arousal

Sexual arousal is defined as an emotional response and increased attention to a specific stimulus, followed by a corresponding genital response (Akerman, 2012). There are, however, several challenges to defining *deviant* sexual arousal, as it depends on defining the limits of normative behavior. Thus, definitions of deviance are subjective and dependent upon what is commonly accepted by society during the time which the definition is decided upon (Best, 2011). As such, definitions of what is and is not deviant inevitably differ across time periods and locations. For example, masturbation, a sexual behavior that is now common and generally accepted in most cultures, was once declared an act of sexual deviance (Hewitt, 2002). In fact, any form of "recreational" sex, that which could not result in procreation, was once considered sexually deviant (Hewitt, 2001).

Currently, deviant sexual behavior is defined as sexual behavior that violates norms and/or is denounced by society (Bryant, 2011). As mentioned previously, while the commission of a sexual offense represents deviant *behavior,* it is not sufficient evidence of deviant sexual *arousal,* though deviant sexual arousal is certainly one of many possible motivations for having committed a sex offense. There are several possible reasons why an individual may commit a sexual offense, and correspondingly, not all sexual offenders exhibit deviant sexual arousal (First & Halon, 2008).

Ultimately, an individual is said to exhibit deviant sexual *arousa*l when the stimuli that elicit sexual interest for that individual are illegal, highly unusual in that they differ from societal norms, or cause (undesired) harm to others when expressed behaviorally (Hanson & Morton-Bourgon, 2005; Seto, 2001). When this aforementioned sexual arousal to non-normative stimuli creates distress for the individual or others, the sexual deviance warrants a diagnosis of a paraphilia, according to the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (*DSM-IV TR*; APA, 2004). More recently, in an attempt to remove some of the stigma from sexual interests that are unusual, the *DSM-5* has slightly altered the paraphilia diagnostic categories. It now requires recurrent sexually arousing fantasies, urges, or behaviors that fall into one of seven categories (voyeurism, exhibitionism, frotteurism, masochism, sexual sadism, pedophilia, fetishism) and that either a) the individual acts on the urges with a non-consenting person or b) the urges/fantasies cause distress and impairment to the individual (APA, 2013).

As mentioned previously, conviction for a sexual offense, while a marker of deviant behavior, is not sufficient evidence of deviant sexual arousal, nor is it sufficient to meet criteria for a paraphilia diagnosis*,* as research has shown that not all sexual offenders possess deviant sexual arousal (First & Halon, 2008; Hanson & Bussiére, 1998; Hanson & Morton Bourgon,

2005). In fact, only a minority of sex offenders, approximately 30%, exhibit deviant sexual arousal patterns (Harris, 2014). Among pedophiles, for example, the literature suggests that approximately 40-50% of these sexual offenders with child victims exhibit normative sexual arousal patterns and thus do *not* possess deviant sexual arousal (Seto, 2009).

Given that the majority of sex offenders do not exhibit deviant sexual arousal, it follows that there are several other reasons that an individual may commit a sexual offense. The nonpedophilic child molesters described above, for instance, may have offended against a child due to hypersexuality, lack of available sexual opportunities with preferred adult partners, or possession of an impulse control disorder (Seto, 2008). A perpetrator of rape, while preferring consensual sex, may have misinterpreted his nonconsenting partner's sexual interest (Hanson & Bussiére, 1998). It should also be noted that among normative, nonsexual offending individuals, approximately 10% have deviant sexual arousal/interests (Harris, 2014). Therefore, it is clear that perpetration of a sexual offense is not sufficient to establish a pattern of deviant sexual arousal, nor is the absence of a sexual offense sufficient to establish a lack of deviant sexual interests.

In sum, the sex offender population includes not only those with deviant arousal patterns, but also lower risk offenders, juvenile offenders, statutory rape offenders, and those whose offending patterns are situational (Finkelhor, 2009). As mentioned above, the perpetration of a sexual crime is not sufficient to infer deviant sexual interest and arousal, since situational factors may have been the primary contributing factor. Correspondingly, a lack of subsequent sexual reoffending is not proof that the offender does *not* possess deviant sexual arousal patterns, as behavioral expression of deviant sexual arousal might not take place due to lack of opportunity and/or adherence to social norms (Abel et al., 2004; Quinsey, Rice, Harris, & Reid, 1993). Since

behavioral observations are not sufficient to determine whether deviant sexual arousal is present, it has become necessary that a valid and reliable measure of deviant sexual arousal be established in order to evaluate future risk, as well as treatment-related progress (Akerman, 2008; Akerman & Beech, 2012). However, in addition to the problems inherent in *defining* deviant sexual arousal, there are also several challenges to *assessing* deviant sexual arousal, as the current measures have been criticized for being invasive, complicated and expensive to administer, and vulnerable to feigning.

Despite the difficulty surrounding the definition and assessment of deviant sexual arousal, clinicians are routinely asked to make decisions about whether or not a paraphilia diagnosis is warranted. These decisions have serious consequences, as they influence sentencing and treatment recommendations, as well as recommendations for sexually violent predator (SVP) civil commitment. To assist clinicians and evaluators in arriving at the most accurate evaluation of deviant sexual arousal (DSA), several methods have been proposed and investigated in the literature.

The Assessment of Deviant Sexual Arousal

Self-Report Measures

One such method of assessing DSA is to administer self-report questionnaires. Some of the most commonly used psychometric measures of sexual deviance include the Multiphasic Sex Inventory (MSI; Nichols & Molinder, 1984), the Multidimensional Inventory of Development, Sex, and Aggression (MIDSA; Knight, 2008); the Wilson Sex Fantasy Questionnaire (WSFQ; Wilson, 1978), My Private Interest Measure (MPI; Farren & Barnett), and the Sexual Fantasy Questionnaire (SFQ; Gray, Watt, Hassan, & MacCulloch, 2003).

The Multiphasic Sex Inventory. The most commonly used self-report measure of deviant sexual arousal is the MSI (Nichols & Molinder, 1984), in which offenders are asked to respond to items regarding deviant sexual activities, behaviors, and thoughts, and includes several validity scales, including a sexual social desirability scale (Akerman & Beech, 2012). Research has demonstrated that the MSI exhibits moderate to good internal validity (Akerman & Beech, 2012; Kalichman, Henderson, Shealy, & Dwyer, 1992) and that it can reliably distinguish offender subtypes (Kalichman, Szymanowski, McKee, Taylor, & Craig, 1989).

While it is possible for offenders to deny deviant sexual interests on self-report scales in an attempt to present themselves in a more socially appropriate manner, the MSI includes both social desirability and lie scales to detect when this is the case. The MSI also contains repeated and similar items to check for consistency. A criticism of this measure is that the items are listed in both past and present tense (i.e., "I have," "Sometimes I have," and "I am"), which some researchers propose may limit the ability to detect treatment-related changes (Akerman & Beech, 2012). Nonetheless, Nichols and Molinder (1984) discovered significant differences in the MSI scores between untreated sex offenders and those who successfully completed treatment. Schlank (1995) demonstrated that the MSI was better at addressing sexual deviance and more effective in distinguishing offender subtypes than the Minnesota Multiphasic Personality Inventory (MMPI; Kalichman, Szymanowski, McKee, Taylor, & Craig, 1989). However, no predictive validity studies have been conducted to examine the relationship between treatment-related changes in MSI score and sexual recidivism.

The MSI-II (Nichols & Molinder, 1996) was later developed to expand and improve upon the original self-report measure. This amended edition includes 560 true/false items and provides unique forms for females and adolescents. It asks questions about sexual interests in child

molestation, rape, exhibitionism, voyeurism, pornography, bondage and sadism/masochism (Tong, 2007). The psychometric properties are available in a published handbook, and the creators report improved reliability and validity for the 40 scales on this measure. Little research has been published on this measure. However, one study comparing cut-off scores for sexual deviance among 51 adult males accused of child molestation using Penile Plethysmography (PPG), the Abel Screen, and the MSI-II found that those categorized as sexually deviant and non-deviant scored similarly on the three measures, lending support for the MSI-II as a corroborating measure of deviant sexual arousal among child molesters (Tong, 2007).

Wilson Sex Fantasy Questionnaire. Another self-report measure of deviant sexual arousal is the WSFQ (Wilson, 1978). The WSFQ is comprised of 40 items aimed to assess exploratory, intimate, impersonal, and sadomasochistic sexual fantasies. Test developers have made this measure available online, as well as the scoring guidelines, norms, and the administration manual. Gosselin and Wilson (1980) found that individuals in the general public who had sexually deviant interests exhibited significantly higher scores on the impersonal and sadomasochistic fantasy subscales than non-deviant test takers. However, there has been little research conducted to investigate the measure outside of the laboratory of its developers. Additionally, Baumgartner, Scalora, and Huss (2002) note that while the WSFQ addresses rape, violent sex, and other fantasies related to sexual offending, the measure does not contain subscales addressing sexual fantasies involving children, and is thus limited in its ability to differentiate between child molesters and non-offenders. Nonetheless, Baumgartner and colleagues (2002) found that among their sample of offenders and controls, child molesters exhibited higher overall scores, exploratory scores, and intimacy scores than non-sexual offenders. However, both child molesters and non-sexual offenders reported fewer fantasies than

control examinees, limiting the usefulness of the measure. Further, the relationship between WSFQ scores and sexual recidivism has not been examined (Baumgartner et al., 2002).

My Private Interest Measure. The My Private Interest Measure (Farren & Barnett, in press) is a 54-item self-report measure which includes four subscales: sexual preference for children, obsession with sex, preference for violent and/or humiliating sex, and other offense-related sexual interests (Akerman & Beech, 2012). This measure appears promising, as it addresses a wider variety of deviant sexual interests than its predecessors; however, to date, there have been no published studies examining the construct or predictive validity of this measure. Further, Akerman and Beech (2012) note the items on this measure are worded only in the present tense and that this may be problematic in terms of using the measure to detect treatment-related change. Two more recently developed measures of deviant sexual arousal are the Multidimensional Inventory of Development, Sex, and Aggression (MIDSA; Knight, 2008) and the Sexual Fantasy Questionnaire (SFQ; Gray, Watt, Hassan, & MacCulloch, 2003).

Multidimensional Inventory of Development, Sex, and Aggression. The MIDSA is a computerized self-report measure that is sensitive to random responding by measuring time to respond and is individualized, in that the computer software makes determinations about which items to provide, depending on the test taker's previous answers (Akerman & Beech, 2012). It was developed to aid clinicians in identifying treatment targets for sexual offenders and is most accurate in assessing in the domains of antisocial behavior/values, developmental and family history, deviant sexual interests/arousal, education/employment history, relationship and sexual history, personality traits, substance abuse, and use of sexually arousing materials (MIDSA Manual, 2011). The developers note, however, that the measure has not yet been evaluated for its

predictive accuracy, and is thus not currently helpful as a recidivism risk assessment instrument (MIDSA Manual, 2011).

Sexual Fantasy Questionnaire. Lastly, the SFQ (Gray, Watt, Hassan, & MacCulloch, 2003) is a self-report measure initially used to assess the rates of sadistic sexual fantasy among a non-offending sample. Maile and Jeglic (2009) administered a modified version of the SFQ, the Sexual Interest Questionnaire (SIQ) to university undergraduates to examine the frequency and type of deviant sexual interests possessed by this non-offending sample. Consistent with prior research, the majority of the sample (86%) endorsed deviant sexual interests (Maile & Jeglic, 2009). Further, the deviant sexual fantasies exhibited by the undergraduates were quite diverse and included sadistic, masochistic, voyeuristic, exhibitionistic, rape-related, bestiality-related, and pedophilic interests (Maile & Jeglic, 2009). The measure shows promise in that it was able to detect a large prevalence of wide-ranging deviant sexual interests in a community sample. However, research on this measure is still in the preliminary stages, and its predictive abilities have not yet been empirically examined.

Overall, self-report measures of DSA can be useful in that they provide a very inexpensive and relatively quick method of assessing sexual interests. Research has shown self-reported DSA to strongly correlate with physiological indicators of arousal, while overcoming the ethical limitations of physiological measures, which are considered to be very invasive and will be discussed in greater detail below (Akerman & Beech, 2012; Tong, 2007). Among the self-report methods, questionnaires may provide the most amount of information about deviant sexual arousal, as compared to other self-report methods, such as a direct interview (Akerman & Beech, 2012).

However, self-report is vulnerable to social desirability, whereby respondents may minimize or lie about their deviant sexual interests. Further, self-report requires that the respondent have insight into their own sexual interests and arousal, and some individuals may lack insight, thus further limiting the usefulness of self-report measures (Akerman & Beech, 2012; Crooks Rostill-Brookes, Beech, & Bickley, 2009). Lastly, there is a dearth of predictive validity studies aimed at examining the correlation between scores on sexual deviance self-report measures and recidivism. In order to overcome some of the social desirability obstacles and provide more reliable assessment of deviant sexual arousal, researchers have developed physiological measures to detect involuntary bodily changes during arousal.

Penile Plethysmography

One of the most reliable, well established, and widely used measures of DSA is the phallometric assessment, also known as penile plethysmograpy (PPG; Akerman & Beech, 2012; Crooks, Rostill-Brookes, Beech, & Bickley, 2009; Malcom, Andrews, & Quinsey, 1993). Penile Plethysmography is a measure of genital arousal, specifically, penile circumference or volume, in the presence of sexual stimuli (Akerman & Beech, 2012; Seto, Lalumiere, & Blanchard, 2000). This method requires that a sensor device be placed around the offender's penis. Most commonly, a mercury-in-rubber strain gauge is the sensor device used, and it is placed around the shaft of the penis (Seto, 2001). The offender is then presented with either visual and/or auditory sexual stimuli, while the device on the individual's penis takes measurements of blood flow in the presence of these stimuli. The stimuli might include, for example, pictures of nude and/or sexually suggestive individuals of various ages or audiotapes of consenting or nonconsenting sexual encounters (Kercber, 1993). The measurement data is then interpreted by a trained evaluator to determine the respondent's level of sexual arousal to the stimuli. Increases in

genital arousal are interpreted in terms of relative responding, that is, in comparison to non-deviant responding for the individual being assessed, and the magnitude of the response is interpreted as the valence of sexual interest in the auditory or visual stimuli presented (Seto, 2001).

Phallometry has been considered the gold standard of deviant sexual arousal assessment because it is highly specific to sexual responding (Seto, 2001). This degree of specificity has not been attained using more general physiological measures, such as heart rate, breathing patterns, and galvanic skin response, which are commonly administered along with the PPG but are affected by a variety of confounding factors (e.g., general arousal; Barker & Howell, 1992; Freund, 1977). Further, this method may overcome some of the main complications inherent to the use of self-report measures, such as the ability of offenders to minimize or conceal their deviant sexual interests and lack of insight (Seto, 2001).

The reliability and validity of PPG as a measure of DSA has also been generally well-established in the literature (Akerman & Beech, 2012; Crooks, et al., 2009; Malcom et al., 1993). Researchers have detected significant differences in genital responding to deviant stimuli between child molesters and general nonsexual offenders, between child molesters and non-offending samples, and between rapists and non-rapists (Freund & Watson, 1991; Lalumiére & Quinsey, 1994; Seto et al., 2000). More specifically, those who offend against children respond more to child-related stimuli than to adult-related stimuli, when compared to other types of offenders and non-offending populations (Seto et al., 2000). Conversely, non-offending samples exhibit the strongest responses to adult stimuli, as compared to adolescent and child stimuli (Barbaree & Marshall, 1989).

The PPG has also been shown to accurately discriminate among sex offender subtypes (Blanchard, Klassen, Dickey, Kuban, & Blak, 2001). Malcom, Andrews, and Quinsey (1993) found that among 172 convicted sex offenders, subjects responded at the highest levels to visual stimuli depicting models most consistent (e.g., gender, age) with their most recent victim. For instance, child molesters exhibited a greater preference for children compared to other sex offenders, and those child molesters with prepubescent victims exhibited the greatest preference for younger models (Malcom et al., 1993).

Perhaps the most desirable quality of a measure of deviant sexual arousal is its relationship to future reoffending, as the ultimate intent is to use the measure to conduct sound recidivism risk assessment. The accuracy of PPG testing in predicting recidivism has been investigated empirically. The extant literature suggests that there is a small to moderate correlation between DSA as measured by phallometric testing and future sexual recidivism. Barbaree and Marshall (1988) demonstrated that DSA as measured by PPG was significantly related to recidivism ($r = .38$) among a small sample of 35 untreated child molesters. Since then, other studies have replicated the finding that sex offenders displaying greater sexual deviance in phallometric testing are more likely to commit a new sex offense upon release (Quinsey et al., 1995; Serin et al., 2001).

Hanson and Bussiére's (1998) meta-analysis demonstrated that sexual interest in children, as measured by the PPG, was the best predictor of sexual recidivism out of all static and dynamic variables examined. However, sexual interest in rape, as measured by PPG, was not significantly related to recidivism ($r = .05$; Hanson & Bussiére, 1998; Marshall, 2006). Malcom and colleagues (1993) found sexual recidivists to exhibit significantly more deviant age

preferences than non-recidivists; however, they failed to detect a statistically significant difference between sexual recidivists and nonsexual recidivists (Malcom et al., 1993).

A valuable aspect of the PPG test is that phallometric data may add incremental predictive accuracy *beyond* what can be gleaned from self-report, even when offenders have already admitted to deviant sexual arousal (Seto, 2001). That is, the *strength* of the deviant interests is thought to be directly related to risk for recidivism, such that an offender with a higher level of responding on the PPG in the presence of deviant stimuli is more likely to recidivate than an offender with a lower level of responding, even when both examinees admit to having a deviant sexual interest on a self-report measure (Seto, 2001).

It has been suggested that while PPG results alone exhibits moderate predictive accuracy, the combination of PPG with other predictive measures may increase accuracy (Marshall, 2006). Quinsey, Rice, and Harris (1995) found that the PPG, combined with psychopathy ratings and previous criminal histories together were able to predict sexual recidivism among 219 sexual offenders released from a maximum security psychiatric facility with 77% correct decisions ($p <$ .001). The relationships between sexual deviance and recidivism and between psychopathy and recidivism were small, but together demonstrated an additive relationship with recidivism, suggesting predictive accuracy may be enhanced when psychopathy measures are added. Similarly, Serin and colleagues (2001) found improved accuracy in predicting sexual recidivism when high Psychopathy Checklist- Revised (PCL-R; Hare, 1991) scores were combined with PPG scores. Those labeled highly sexually deviant (according to PPG) with high PCL-R scores recidivated earlier and more frequently than highly sexually deviant offenders with low PCL-R scores (Serin, et al., 2001). Overall, it appears that there is a small but significant relationship between PPG testing and recidivism, and that it may be most effective when considered in

conjunction with other predictive factors, such as criminal history and psychopathy (Marshall, 2006; Quinsey, Rice, & Harris, 1995).

Despite its relative success, phallometric testing has elicited several significant ethical and pragmatic criticisms. First, the PPG methodology itself has been criticized. The PPG is considered to be a very invasive process, as it requires the subject to remove his clothing in order to secure a sensor device to the penis, so that genital responses may be recorded by computer technology (Seto, 2001). Further, many have taken issue with giving so much weight to genital arousal as measured by erection as a primary indicator of sexual arousal to the point of neglecting a variety of other physical, physiological, cognitive, emotional, and behavioral aspects of sexual arousal (Kalmus & Beech, 2005; Laws, 2009; Singer 1984).

Research has also demonstrated a lack of standardization in the administration procedures and statistical analysis of phallometry (Laws, 2009; Marshall, 2005; Seto, 2001). This is due, in large part, to a lack of commonly accepted guidelines for conducting PPG methodology (Akerman & Beech, 2012; Laws, 2003; Odeshoo, 2004; Seto, 2001). For example, researchers have found vast differences in the types of gauges utilized, the type and number of stimuli presented, examinee instructions, measurement intervals, efforts to assess/prevent faking, and the type of data analysis employed in PPG administrations (O'Donahue & Letourneau, 1992). Without the creation of uniform administration and analytic procedures to ensure that the measure is used in a reliable manner across examinees, phallometric testing does not meet the standards required for psychological tests (American Psychological Association, 1999).

Another criticism of the PPG is that it is extremely costly and time consuming to administer. Recent estimates suggest that administration of the PPG can cost up to $1,000 per session (Odeshoo, 2004).

Lastly, PPG examiners have to be specifically trained to administer and analyze PPG results. Given the cost, time, and resources required to do so, there is also evidence suggesting that there has been a lack of standardization in training. A survey of PPG assessment centers demonstrated large variations in the degree of training and experience of PPG clinicians, ranging from 6 months to 25 years (Howes, 1995). Further, 18% of evaluators reported having no advanced training at all (Howes, 1995). Moreover, researchers note that the standardization procedures necessary to improve the consistency of administration and interpretation across evaluators would likely require even more time and cost (Abel, Huffman, Warberg, & Holland, 1998; Akerman & Beech, 2012).

It is worth noting that some efforts have been made to standardize the PPG administration procedures. For example, guidelines for the utilization of PPG testing have been published in the *Psychiatric Bulletin* (Fuller, 1995). The Association for the Treatment of Sexual Abusers (ATSA) has also published guidelines for conducting PPG evaluations in *Practice Standards and Guidelines for Members of ATSA* (ATSA, 2004). However, the degree to which evaluators abide by these guidelines is unknown (Howes, 1995; Prentky & Burgess, 2000).

Despite the fact that the PPG has been thought to overcome confounding factors like social desirability, recent evidence suggests that the penile plethysmography can in fact be faked by the examinee (Abel et al., 2004; Akerman & Beech, 2012; Kalmus & Beech, 2005). More specifically, research has demonstrated that offenders may look away from the stimuli, generate and focus on their own visual imagery, or otherwise distract themselves, in an attempt to increase genital responding to normative stimuli and/or to decrease genital responding to deviant stimuli

(Abel et al., 2004; Kalmus & Beech, 2005). Seto (2001) has noted that the placement of a camera aimed at the upper body of the examinee helps to decrease attempts to look away or alter the placement of the sensory device.

Another method of distorting the test results is referred to as "pumping," in which examinees employ muscle contractions to increase erectile response to sexually normative stimuli (Abel et al., 2004). Still other examinees may respond very little or not at all to both deviant and normative stimuli. In fact, research has demonstrated that as many as 20-30% of examinees exhibit levels of responding that are too low in the presence of both deviant and normative stimuli to be interpreted. This is especially problematic, as these individuals produce invalid and uninterpretable data, which provides little to no valid information about the sexual arousal patterns for those individuals (Abel et al., 2004; Serin, Malcom, Khanna, & Barbaree, 1994).

Finally, the use of PPG as a measure of DSA has elicited numerous ethical criticisms. First, the presentation of pictures of actual children as stimuli may be considered unethical, as the children cannot provide informed consent (Beech et al., 2008; Seto, 2001). Furthermore, the presentation of nude or suggestive photos of children which have been seized by police, an option that has been considered in the past, may also be unethical, as it may constitute revictimization of the individuals in the photographs (Abel, Huffman, Warberg, & Holland, 1998). Due to several ethical concerns, the PPG remains unavailable in many European countries (Babchishin, Nunes, & Hermann, 2013).

To combat some of these ethical issues, Seto (2001) has highlighted that audiotaped, rather than visual, stimuli can be used, in order to avoid the presentation of child photographs, which would essentially be child pornography. Further, specialized software is available that

allows for the creation of digital composite images which appear highly realistic but do not portray actual individuals or require the use of real human photographs (Seto, 2001). However, the generalizability of auditory stimuli and/or composite images to images of visual stimuli depicting actual individuals remains unknown.

Questions have also been raised about the ethicality of presenting deviant sexual stimuli, which may include sexual images of children, sexual violence, and/or nonconsensual sex, as part of a risk assessment protocol for first-time offenders or to adolescents (Abel et al., 2004; Seto, 2001). Some critics of this method have suggested that the presentation of these stimuli could potentially have harmful effects on first-time offenders and adolescents. More specifically, some have suggested that since adolescents' sexuality and sexual arousal is malleable, presentation of sexually deviant material may have lasting effects on their cognitive development (Worling, 2011). A counterargument is that adolescent offenders may already be exposed to potentially "harmful" sexually themed material in group therapy sessions, readings, and videos (Seto, 2001). In any case, Seto (2001) has explained that comprehensive debriefings should occur after completion of the phallometric testing, so as to combat any harmful effects of the testing. Further, Seto (2008) cautions that phallometric testing, especially with regard to adolescents, is only indicated for those who are known to have committed a sexual offense and thus have already engaged in criminal sexual behavior, in order to assess their future risk.

Apart from concerns about the potential to cause harm, there is also a lack of research examining the efficacy of PPG methodology with adolescents (Abel et al., 2004). This is an important group to examine, because it is quite common for sex offenders' deviant sexual arousal and resulting behaviors to begin when they are under 18 years old (Abel et al., 2004). The few studies that have been conducted in this area, however, lend support to the efficacy of

phallometric testing, with both auditory and visual stimuli, among adolescent offenders (Becker, Hunter, Goodwin, Kaplan, & Martinez, 1992; Seto, Lalumière, & Blanchard, 2000).

Overall, the PPG is currently considered the best measure of sexual deviance, because it is highly specific to sexual, as opposed to general arousal, its validity has been well-established in the literature, and despite the ability for some individuals to distort or suppress their responses, it remains less vulnerable to impression management and social desirability than self-report measures. However, its lack of standardization, questionable reliability, and somewhat inconsistent predictive accuracy, combined with its invasiveness and ethical concerns, suggests there is still much room for improvement in the assessment of sexual deviance.

Attention-Based Assessment

Given that physiological measures of DSA are replete with ethical and financial constraints, and self-report measures, although cost-effective and easy to administer, are vulnerable to social desirability and lack empirical support of their predictive accuracy, recent research has focused on novel methods of measuring DSA that are less invasive, less costly, and more ethically sound. An area of much recent attention is the examination of attention-based paradigms for measuring DSA including visual reaction time, attentional eye blink, implicit association task, and modified Stroop tasks. The theoretical basis for attention-based measures is that an information processing delay can be induced by presentation of a sexual stimulus. The sexual stimulus creates a demand in the individual's resources, which takes away from the available resources to process the relevant task information and therefore causes increased difficulty and delay in responding (Ó Ciardha, 2011).

Visual reaction time methods. One such attention-based method uses reaction time measurements to assess deviant sexual arousal. In visual reaction time methods, computerized

visual stimuli are shown to respondents, and the respondents are asked to make ratings for each picture in terms of attractiveness or preference and then press a key to move to the next image (Osborn, Abel, & Warberg, 1995; Seto, 2001). The amount of time spent viewing each picture, referred to as the visual reaction time (VRT), is recorded. Longer viewing times for images of preferred sexual objects are used as an indicator of sexual interest in image content. In other words, longer response times to deviant stimuli indicate an attentional bias toward the deviant image content (Ó Ciardha, 2011).

Among non-offending samples, research has shown visual reaction time measures to be highly correlated with self-reported sexual interest, as well as phallometric responding (Abel, Rouleau, Lawry, Barrett, & Camp, 1990; Seto, 2001; Tong, 2007). Abel and colleagues (1990) first employed a visual reaction time method, using visual images of 80 men and women of various ages, among heterosexual and homosexual men. The results demonstrated significantly longer reaction times for images of individuals that were gender-consistent with the examinees' sexual preference than for images of the opposite gender (Abel et al., 1990).

Research employing visual reaction time methods among known offender samples has also shown promise. Abel and colleagues (1990) compared the reaction times of non-offenders to admitted child molesters with male, female, child, and adolescent victims. They used visual reaction times to categorize offenders into one of four categories: interest in child males (under age 14), adolescent males (14-17 years old), child females, or adolescent females. Results indicated that visual reaction times were most accurate in categorizing those with interest in adolescent males and females, moderate accuracy in categorizing those with interest in male children, and low accuracy for the female children (Abel et al., 1990). There is also some evidence to suggest VRT may perform equally to, if not better, than PPG measures of DSA, as

VRT has been shown to correlate significantly with victim choice for female children, male children, and adolescent males, while PPG results correlated with only two categories, male children and female adolescent victims. Further, correlations reached significance only when nonresponders, those with little to no erectile change (defined as those with less than 30% of a full erection for any stimuli), were included in the analysis (Abel et al., 1998).

The Abel Assessment for Sexual Interest (AASI; Abel et al., 2004) includes an evaluation of subjective ratings of sexual interest, VRTs to deviant sexual stimuli, and responses to a questionnaire concerning criminal history and sexual interest. The literature has demonstrated that upon administering the AASI, the VRT results accurately differentiate between admitted child molesters and non-offenders, such that adolescents with greater VRTs for child-related stimuli were significantly more likely to be child molesters than those with shorter VRTs. Further, among child molesters, there was a significant positive correlation between VRTs in response to child-related stimuli and number of child victims (Abel et al., 2004).

These results have not, however, been consistently replicated in other research laboratories or with other populations. For instance, Smith and Fischer (1999) administered the Abel Assessment for Interest in Paraphilias (AAIP), a method which, similar to the AASI, utilizes self-report, VRT, and a questionnaire to a sample of 81 adolescent offenders and non-offenders. No statistically significant differences in VRTs to deviant stimuli were found between offenders and non-offenders, and their results did not demonstrate sufficient test-retest reliability (Smith & Fischer, 1999). In 2008, Stinson and Becker employed the AASI among a sample of known sexual offenders and also did not find a statistically significant correlation between VRTs to deviant stimuli and stimulus-consistent deviant behaviors, such as fetishism, frotteruism, or zoophilia (Stinson & Becker, 2008).

Critics of VRT as a measure of DSA claim that it may have little clinical utility, since there is no "correct" answer for each trial and subjects are not encouraged, explicitly or implicitly, to work as quickly or as accurately as possible (Ó Ciardha, 2011). In other attention-based models, which will be discussed in greater detail below, there is a correct answer for each trial, and respondents must balance speed and accuracy. Response latencies are therefore rather quick, and interference effects are typically measured in milliseconds. Thus, many of the attention-based methods are thought to be robust against efforts to purposely distort response profiles. With VRT methods, however, there is no correct answer and response latencies may therefore be longer in general than that of other attention-based tasks. Further, VRT methods may be vulnerable to conscious distortion if a respondent is aware that speed is being measured (Ó Ciardha, 2011). Therefore, it is unknown to what degree the time required to select a choice accurately represents deviant sexual interests. Consequently, researchers have noted that the theoretical foundation for the VRT method is lacking, as the research has not been consistent and it is unclear what specifically causes the increased viewing time for sexually attractive stimuli (Akerman & Beech, 2012; Imhoff et al., 2010).

In addition to the lack of available published research regarding this method of assessing sexual deviance, there are some other limitations to using this method as well. First, concerns have been raised about the use of deception involved in VRT methodology (Akerman & Beech, 2012). Specifically, examinees are not informed that the length of viewing time, rather than stimulus ratings, is being used to assess DSA (Akerman & Beech, 2012). Whereas some have taken issue with the deceptiveness of the measure, still others have suggested that the purpose may actually be too transparent to offenders and thus has the potential to be faked (Kalmus & Beech, 2005; Seto; 2001; Smith & Fischer, 1999).

Additionally, VRT methods may elicit some of the same ethical criticisms that have been raised with regard to penile plethysmography. The stimuli commonly utilized in the VRT method depict images of individuals of various ages, including children. Thus, some may take issue with the presentation of images of actual children, as some argue it is revictimization, as mentioned above. However, this measure assesses viewing time rather than penile circumference, and erection responses are not necessary. Thus, there may be less of a need to present nude and/or sexually explicit photographs, although it is likely that stronger manipulations (e.g., nude as opposed to clothed) produce stronger effects. For example, the stimuli used by Abel and colleagues (1998) in their VRT study included individuals in bathing suits against a blue background. Given that VRT methods may not require the use of nude individuals, it may be viewed as more ethically sound than the PPG.

Overall, the evidence accumulated thus far appears to lend some support for the usefulness of the VRT method in the assessment of deviant sexual arousal, and it is indeed less invasive, less costly, and may be more ethical to administer than the PPG. However, considerably more research needs to be conducted, across laboratories and with various offender populations, including those who deny involvement in sexual offenses, to determine conclusively the soundness of the measure (Crooks et al., 2009). Further, future research should assess the accuracy of this method regarding the prediction of sexual recidivism.

Choice reaction time task. The choice reaction time task (CRT) is another attention-based method of assessing DSA, which has also been referred to as an attentional blink (AB) paradigm or dot-probe task. It is considered by some to be a "true" reaction time task, as respondents are asked to identify the location of a dot on an image, and there is one correct answer (Ó Ciardha, 2011). This methodology involves the rapid serial visual presentation

(RSVP) of computerized stimuli, some of which are sexually deviant (targets) and others that are not (probes). The content of the deviant images can contain people, either clothed or nude, when utilized for the purpose of assessing presence of DSA. The respondent's response time is measured and delayed response latencies are interpreted as an indicator of bias toward sexual stimuli (Akerman & Beech, 2012; Kalmus & Beech, 2005). The theory behind the methodology is that increased attention to sexually arousing images, and thus distraction from the task (interfering), results in longer latencies for the sexually arousing images. Empirical investigations of this assessment method, however, have yielded mixed results.

Wright and Adams (1994) investigated the ability of the CRT to detect preferred gender interest among a non-offender sample of eighty community participants (20 heterosexual men, 20 heterosexual women, 20 homosexual men, and 20 homosexual women). The researchers measured the response latencies to stimuli that fell into three categories: nude men, nude women, and neutral stimuli. The results revealed that 95% of participants scored as predicted by their self-reported sexual preferences on the CRT task, with longer response latencies on the trials that depicted preferred-gender images, as compared to non-preferred gender images and neutral images, thus lending support for the interference effects resulting from sexually salient images on a cognitive task (Wright & Adams, 1994). In an expansion of their initial pilot study, Wright and Adams (1999) replicated their CRT task with 80 additional community participants but added trials in which the male and female stimulus images were clothed. As hypothesized, respondents exhibited the longest response latencies on trials where the stimuli elicited the greatest level of sexual interest (e.g., nude, preferred gender images). Further, the trials with clothed preferred gender images elicited longer response latencies than the neutral stimuli but shorter latencies than the nude preferred gender images (Wright & Adams, 1999).

Some studies have found that the CRT task is also able to accurately identify group differences when applied to sexual offender populations (Giotakos, 2006; Mokros, Dombert, Osterheider, Zappala, & Santtila, 2010). Mokros and colleagues (2010) administered a CRT task to a sample of 21 sexual offenders against children and 21 nonsexual offenders mandated to inpatient treatment at a forensic psychiatric hospital for a criminal offense. The results revealed that the child molester sample had longer response latencies to identify the position of the dot on trials where the stimulus contained an image of a child than for adult stimuli. Conversely, nonsexual general offenders exhibited longer response latencies for the adult stimuli than the child stimuli. Further, the response latencies for both groups were longer for the nude stimuli than the clothed stimuli that corresponded to their sexual interest, either child or adult (Mokros et al., 2010). Giotakos (2006) similarly examined the ability of the CRT task to assess deviant sexual interest using semi-nude images of males and females of various ages among a sample of 31 rapists, 27 child molesters, 53 control males, and 24 control females. The results indicated that extra familial child molesters had the longest response latencies for images of girls, while intra-familial child molesters had the longest response latencies for adolescent females, and rapists and control males had the longest response latency for adult females (Giotakos, 2006).

However, there were not significant differences between the control males' profiles and the profiles of the rapists. Moreover, the authors noted that the rapists' profiles were similar to that of the control males', which is hugely problematic for trying to distinguish between groups. Furthermore, other research groups have been unable to replicate the results supporting the efficacy of CRT in detecting sexual interests. Gress, Anderson, and Laws (2013) for example, found that while a similar CRT task was able to detect significantly different response latencies between youth non-sexual offenders and a university sample, it was unable to distinguish the

adult sex offenders from the youth non-sexual offenders or the adult sex offenders from the community sample.

Beech and colleagues (2008) employed a slightly different RSVP measure to examine attentional blink effects among a prison sample including 35 known intrafamilial and extrafamilial sex offenders, as well as a comparison group of 17 nonsexual offenders incarcerated for other crimes. In this study, a series of visual images were presented to examinees, and they were asked to respond to memory questions (e.g., "Did you see a train or a chair?") about two target images. The theory behind the methodology is that in general, memory for the second target is less accurate when it is presented shortly after the first target, but that the amount of error will differ based on the content of the first image and the examinee's attention to that image. The results indicated that child molesters were significantly less accurate than the comparison offender group to report the second target when the first target image was a picture of a child. Contrary to their hypothesis, however, they did not find that extrafamilial (more likely pedophilic with deviant sexual arousal patterns) child molesters made more errors than intrafamilial (more likely to exhibit normative, adult sexual preference) child molesters (Beech et al., 2008).

Subsequently, Crooks and colleagues (2009) investigated the use of RSVP with an adolescent sex offender population recruited from inpatient and outpatient sex offender treatment facilities, as compared to a sample of adolescent nonsexual offenders from a Youth Offending Service. Inconsistent with their hypothesis, the researchers did *not* find the adolescent sex offender group to display increased errors in identifying the second target when the first target was a child. Rather, they found that both the sex offenders and nonsexual offenders made more errors in identifying target two when the first target was an animal, as opposed to a child (Crooks

et al., 2009). Thus, the results do not support the efficacy of RSVP measures in discriminating adolescent sexual offenders from non-offending adolescents. It may be the case that adolescents in general, due to immaturity, attend more to the animal images than the human images. The researchers also suggest that the subtlety of the images (i.e., the individuals depicted in the images are clothed) may have prevented the detection of group differences.

Overall, there is a lack of research aimed at investigating the efficacy of RSVP, otherwise known as attentional blink methods, in the assessment of deviant sexual arousal. Further research with larger sample sizes and addressing the validity and test-retest reliability of this method are needed before conclusions can be drawn about its usefulness or lack thereof (Crooks et al., 2009; Kalmus & Beech, 2005). Further, if the usefulness of RSVP in detecting sexual deviance is in fact established, it will then be critical for studies to empirically examine the accuracy of RSVP methods in predicting recidivism, as deviant sexual arousal is a significant risk factor for repeated sexual offending. Lastly, this method also requires the use of images and will thus inherit the same ethical concerns that have been raised with regard to PPG testing, and this must be taken into consideration when evaluating its clinical utility.

The Implicit Association Test. The Implicit Association Test (IAT) is a cognitive measure in which participants are asked to categorize stimuli into categories designated by computer keys. The ease with which a participant categorizes stimuli to a particular key, measured by response time, is interpreted as the strength of the mental association between the stimulus and the category, for the respondent (Ó Ciardha & Gormley, 2012). When utilized as a measure of DSA, researchers assess the response time for a respondent to pair deviant words or images with the sex category. For example, researchers may compare the latency for categorizing children with sex to the time required to pair adult and sex. For this paradigm,

contrary to the other cognitive measures, *shorter* response latencies are interpreted as greater deviant interests, as the relationship between the deviant stimulus and sex is more salient (Ó Ciardha & Gormley, 2012).

In a recent meta-analysis investigating 12 IAT studies from the years 2004 to 2011, the researchers found that IAT measures were able to distinguish sex offenders against children from non-offenders, with a moderate to large effect size (Babchisin et al., 2013). The IAT performed slightly worse (i.e., small to moderate effect sizes) when comparing sex offenders against children to general offenders and rapists than when compared to non-offenders. Given the small number of existing studies that have examined the IAT, there was not a significant correlation between IAT measures and actuarial estimates of recidivism risk (e.g., Static-99; Hanson & Thornton, 2000) and there were not enough available effect sizes to assess the relationship between IAT and PPG testing in the meta-analysis. Overall, studies that have employed the categories sex and not sex have been the most successful in discriminating sex offenders from nonoffenders.

Given that very little research exists on the use of the IAT with sex offenders, there is not literature available discussing convergent validity or recidivism predictive accuracy; however, the results of the meta-analysis suggest that IAT may have clinical utility and is related to self-report, sex offense history, and viewing time measures of DSA (Babchisin et al., 2013).

The emotional Stroop. Most recently, the emotional Stroop, another cognitive interference task, has been proposed as a method of assessing DSA. The Stroop, in its original format, involves presenting participants with a list of words printed in different ink colors and requiring them to state aloud the color of the ink while ignoring the meaning of the word (Stroop, 1935). The amount of time required to color name while inhibiting word meaning, the brain's

default method of processing, is recorded and interpreted as the degree of interference. That is, longer latency periods are interpreted as greater distraction by, or information processing of, the meaning of the word. Since the development of the original Stroop, modified versions of this task, manipulating the themes of the word content, have been examined.

For example, research has consistently shown that respondents with affective disturbances exhibit longer latencies than control group participants when asked to color name emotional words that are relevant to their own pathology (Williams et al., 1996). In fact, this interference effect has been demonstrated among not only those with affective disturbances, but also among those with eating disorders (Channon, Hemsley, & de Silva,1988; Fairburn, Cooper, Cooper, McKenna, & Anastasiades, 1991), alcohol disorders (Sharma, Albery, & Cook, 2001), cannabis abuse and dependence (Field, 2005), and sleep disorders (Spiegelhalder et al., 2010). Therefore, researchers hypothesized that sexual interest may influence attention and information processing in a similar manner to the way depression, anxiety, and eating disorders influence information processing (Wright & Adams, 1999). To date, four studies have examined the Stroop interference effect to examine deviant sexual arousal among sexual offending populations using either a word or pictorial Stroop paradigm, and they will be reviewed in detail below (Ó Ciardha & Gormley, 2012; Price, Beech, Mitchell, & Humphreys, 2013; Price & Hanson, 2007; Smith & Waterman, 2004).

Smith and Waterman (2004) administered the Stroop task to a sample of sexual, violent, and nonviolent offenders, as well as a comparison sample of undergraduate students. Their word list was comprised of aggression-related words, sexual words, positive emotion words, negative emotion words, neutral words, and color words. The researchers hypothesized that "sexually deviant offenders," defined by conviction for at least one index sexual offense, would exhibit

longer latencies for sexually themed words than the neutral words as compared to the other offender and nonoffender groups, as the sexual content would be most relevant to their pathology and would therefore exhibit greater interference effects.

The results, however, revealed that both sexual and violent offenders exhibited significantly longer latencies to color name the sexual words than the undergraduate students. Notably, no statistically significant difference in latencies for sexual words was found between sexual and violent, nonsexual offenders (Smith & Waterman, 2004). Thus, this study failed to demonstrate that the Stroop could reliably distinguish between sexual offenders and nonsexual offenders. The researchers explain that the sexual words selected for their list may have also contained violent themes that were highly salient to the violent offenders, which may help explain the lack of significant findings (Smith & Waterman, 2004).

Alternatively, it may be that sexually themed material is highly salient for the majority of individuals, rather than being unique to sex offenders. Perhaps differences in response latencies may be more readily detected if Stroop performance for deviant words is compared to performance on non-deviant or "normative" sexual words, as opposed to making comparisons between sexual and nonsexual material. Further limitations of this study include the small sample size (*n*=43), which may have limited the ability to detect statistical significance, as well as the administration of several questionnaires prior to completion of the Stroop task, particularly, those related to aggression. This may have created confounding effects by priming respondents and thus affecting their performance on the ensuing Stroop task.

Price and Hanson (2007) replicated and expanded upon Smith and Waterman's (2004) study. They administered the modified Stroop task to 30 sexual offenders (15 rapists and 15 child molesters), 15 violent, nonsexual offenders, 15 non-violent, non-sexual, general offenders, and

15 non-offending community participants to assess DSA as measured by Stroop response latencies. Price and Hanson (2007) also included a second Stroop task, in which newly developed offense specific word lists were administered to a sample of rapists and child molesters. They hypothesized that these novel word lists would be more salient to the offender subtypes and therefore better equipped to distinguish among them. They generated a list of words thought to be of interest to rapists and child molesters and presented them alphabetically to five university student judges who rated them on whether they were representative of the category. At least three judges agreed upon 15 child molester words (e.g., molest, innocent, pee pee) and 21 rape words (e.g., humiliate, screw, beg), which were retained for the final list. Other improvements included the utilization of a larger sample size (n=75) and a reduction in the number of questionnaires administered prior to the Stroop task. Measures were carefully selected so as to limit the effects on participants' moods and included a reading level task, a rape attitude scale, a child molestation attitude scale, and the Static-99.

The results indicated that for Study 1, sexual offenders (child molesters and rapists) exhibited a greater interference effect (i.e., were significantly slower) to color name sexually themed words when compared to the community sample. However, only the child molesters (and not the rapists) were significantly slower than the non-violent offenders, and neither child molesters nor rapists were significantly slower than the violent (nonsexual) offenders (Price & Hanson, 2007). For Study 2, no differences in response latencies were found among the groups. All respondents demonstrated greater interference effects for the child molestation and rape words than the neutral words, and the Stroop task could not effectively distinguish among groups. Moreover, no significant differences between groups were found on the self-report measures regarding rape attitudes scale or sex attitudes, suggesting the groups were

heterogeneous and may have contained individuals with and without DSA patterns. Overall, Price and Hanson's (2007) findings were consistent with that of Smith and Waterman (2004), with small effects and nonsignificant group comparisons, with greater interference for sex offenders compared to community samples, but not significantly different from the other types of offenders.

Price, Beech, Mitchell, and Humphreys (2013) noted that results so far from Stroop investigations with sex offender samples may reflect sex drive or preoccupation with sexual material in general, rather than specific sexual interests. To address this, they generated lists of words thought to be even more specific to offender subgroups; in fact, they recruited 18 workers from charities working with sexual abusers and 27 sexual abusers to aid in the creation of a new word stimulus list. They administered their Stroop to 27 offenders (five exhibitionists, four incest offenders, nine pedophiles, seven rapists, and two who had offended against adult and child victims) and 21 violent (nonsexual) offenders convicted of murder or manslaughter. Sex offenders exhibited slower raw reaction times for emotion-personality descriptor (EPD) words (e.g., angelic, devious, promiscuous) and sexual action (SA) words (e.g., fondle, caress, licking) than the offender and nonoffender controls. However, upon calculating bias scores, sex offenders exhibited only significantly larger EPD Stroop effects than the nonoffenders. There were no other significant differences. Their results are the first to reflect the potential for the Stroop to detect significant differences in Stroop interference effects for sex offenders when compared to other types of offenders.

The pictorial modified Stroop task. Most recently, the Stroop has also been modified using pictorial stimuli, presented in various tints, to detect interference effects (P-MST; Ó Ciardha & Gormley, 2012). Participants are required to inhibit attending to the image content

and instead attend to the color tinted over the image. The response times are recorded and interference effects are interpreted as bias toward picture content. Research has demonstrated Stroop interference effects among both community and sex offender populations. Ó Ciardha and Gormley (2012) found that community participants exhibited significantly longer response times for adult stimuli than child stimuli and on trials containing images consistent with their self-reported sexual orientation than for inconsistent images (Ó Ciardha & Gormley, 2012). Among sex offenders, the researchers again used offense behavior as a comparison measure of DSA; however, they noted that not all offenders would be expected to have DSA and therefore included a self-report measure of sexual interest in children, and also made an effort to identify the sex offenders who would be *most* likely to possess DSA, based on relationship to victims, since research suggests offenders with extrafamilial victims are more likely to have DSA than incest offenders (Ó Ciardha & Gormley, 2012; Seto, Lalumiere, & Kuban, 1999).

Overall, sex offenders did not differ from non-offenders in their response times on the P-MST, and there was no significant interaction between group (control, low deviance, high deviance) and trial type (adult vs. child; Ó Ciardha & Gormley, 2012). Paired samples t-tests revealed control participants were significantly slower on adult trials than on child trials. However, among the sex offender groups, neither the low deviance nor the high deviance offenders exhibited significantly different response times for child stimuli compared to adult stimuli. The P-MST was therefore unable to distinguish among offenders and non-offenders.

Critiques of the extant DSA Stroop literature. There are several methodological limitations to the extant Stroop studies examining DSA. First, the researchers failed to empirically examine the validity of the Stroop task as a measure of sexual arousal before attempting to use it to distinguish among offender subtypes. Rather than comparing Stroop

performance to another validated measure of DSA, the corroborating measure was the existence of an index sexual offense. However, as explained above, sexual offense behavior is not a reliable indicator of deviant sexual arousal, as there are numerous reasons why an individual might commit a sexual offense independent from sexual arousal patterns (First & Halon, 2008). For well-supported measures like phallometric testing, researchers have evaluated its basic validity, and there is empirical evidence demonstrating that offenders with the most deviant sexual histories show deviant sexual interests (Barbaree & Marshal, 1989; Seto, Hanson & Bussiére, 1998). As of yet, no such concurrent validity studies have been conducted for the Stroop. Thus, it remains unclear whether increased latencies on the emotional Stroop test actually reflect increased levels of DSA. It is critical that the measure's validity be empirically evaluated before attempts can be made to use it with sexual offending populations with any degree of confidence.

In the Stroop studies to date, offender and nonoffender groups may have been comprised of individuals both with and without DSA patterns, highlighting the difficulties inherent in operationally defining DSA and selecting an appropriate corroborating measure. Perhaps a more informative comparison would be between Stroop performance and another valid indicator of DSA such as self-report or PPG, as opposed to offense behavior.

CHAPTER 3: Current Study

The extant literature demonstrates that deviant sexual arousal is an important factor to consider in sex offender risk assessment, as it has been shown to be a strong predictor of recidivism. While there are several methods to assess deviant sexual arousal, each has its associated strengths and weaknesses. Recently, a version of the emotional Stroop task has been used as a means to detect DSA. However, the studies to date have not tested its basic ability to detect DSA (concurrent validity), but rather have focused on using the Stroop to distinguish between subgroups of sex offenders and between sex offenders and non offending controls. Comparisons between subgroups may not provide the most useful information regarding the validity of the Stroop task as a measure of DSA. Furthermore, making group comparisons requires dichotomizing variables (e.g., sex offender versus non-offender), which may be problematic for validating the Stroop methodology for at least two reasons. First, offender subtypes (i.e., rapists and child molesters) are themselves heterogeneous categories. They include individuals with and without DSA who may not remain stable with regard to their victim choices. Thus, it may be difficult to use sex offender typology to assess concurrent validity. Second, the practice of dichotomizing variables adds significant error, reduces power to detect significance, and leads to underestimating effect sizes (Cohen, 1983). Therefore, an alternative strategy would be to compare Stroop response latencies (a continuous variable) for deviant words with a corresponding continuous measure of DSA. To this end, performance on the Stroop can be compared to either self-reported sexual deviance and/or another valid measure of sexual deviance (e.g., penile plethysmography).

Furthermore, in order to validate the Stroop as a measure of DSA, a wide variety of sexual interests must be present in the population, ranging from more conservative and

normative interests to the more deviant interests. More specifically, an ideal sample would include individuals at both the high and low ends of the spectrum of DSA, so as to compare levels of deviant sexual arousal to the corresponding degree of interference on the Stroop task. Price and Hanson (2007) reported no significant differences between groups on the rape attitudes scale or on the sex attitudes questionnaire. Thus, it is not surprising that significant differences were not found on the Stroop task. Further, contrary to expectations, those with greater self-reported sexual entitlement showed *less* bias on the Stroop than those with lower levels of self-reported sexual entitlement (Price & Hanson, 2007). Perhaps these self-report questionnaires were not accurately assessing the intended construct, as sexual entitlement may be a distinct construct from sexual deviance. Thus, a future validation study would require a widely accepted and accurate corroborating measure of deviant sexual arousal to which Stroop interference effects can be compared.

An alternative explanation for the unanticipated findings in the two aforementioned studies may be that the offenders felt compelled to conceal their deviant sexual arousal (e.g., social desirability). Thus, future attempts to validate the Stroop measure of deviant sexual arousal might be more successful if a non-offending sample is utilized. There is evidence to suggest that deviant sexual interests are not dichotomous in nature but rather exist on a continuum. Therefore, it would be beneficial to validate the methodology using it among non-offending samples who are not motivated to lie or conceal their interests, so long as there is reason to believe that deviant sexual interests may be present.

In a study of reported sexual fantasies, Hunt (1974) found that "being forced or forcing someone else to have sex" was the 5th most common sexual fantasy among normative samples of both men and women. More recently, Ogas and Gaddam (2013) studied the sexual interests of a

non-offending sample by analyzing millions of internet sexual search histories, mostly from three major browsing sites and found several potentially deviant searches. Upon ranking the top 12 searched terms for the three internet browsers, they found that "young" and "youth" were actually the top sexual searches. Other commonly searched deviant sexual searches that made the top 12 list included "bestiality" and "incest" (Ogas & Goddam, 2013). The internet search analysis also revealed that 12.5% of sexual searches were age-related, with a frequency of 579,780 sexual searches for "young," 512,418 for "preteen," and 50,059 for "youngest." Age 13 was the most commonly searched age-related sexual search. Interestingly, users who searched "incest," "bestiality," and "granny porn" had the most extensive sexual search histories general.

Maile and Jeglic (2009) examined deviant sexual interests among a sample of college undergraduates, and found that deviant sexual fantasies were highly prevalent among this sample, with 86% of participants endorsing some form of deviant sexual fantasy. Further, Maile, Maile, and Jeglic (2010) found that the level of deviant sexual interest varied greatly among college undergraduates, and that there were large numbers of individuals at both ends of the spectrum of normal to deviant sexual fantasy, suggesting that DSA may be normally distributed in the population. A recent survey administered at Yale University asking about students' sexual interests and practices revealed that among the 40 students who participated, 9% reported accepting payment for sex (i.e., prostitution), 3% had engaged in bestiality, and 52% had engaged in consensual pain during sex (Hua, 2013; Moran, 2013). Three students (7.5%) also endorsed incestuous sexual fantasies.

CHAPTER 4: Aims & Hypotheses

The current study aimed to investigate the validity of the Emotional Stroop Task for Deviant Sexual Arousal among a nonoffending sample of individuals with a wide range of sexual interests. There were four main goals of the current study:

1. To identify the rates of self-reported deviant sexual arousal among a non-offending, university undergraduate sample. For our purposes, DSA was defined as endorsement of sexual arousal to stimuli that differ from societal norms and/or have the potential to cause harm if acted on behaviorally (e.g., sexual interest in children, rape, physical injury, bestiality).

2. To use the Emotional Stroop Task for Deviant Sexual Arousal to assess the response latencies to color name sexually deviant words relative to the time required to color name non-deviant sexual words and neutral words among those who endorse deviant sexual interests and those who do not.

3. To evaluate the validity of the Stroop methodology as a measure of deviant sexual arousal by correlating performance on the Stroop task with self-reported levels of deviant sexual arousal among the non-offending sample.

4. To determine whether response latencies for deviant words on the Stroop task can predict self-reported levels of deviant sexual arousal using regression analyses.

5. To assess the concurrent validity of the chosen self-report measure by comparing the SFQ, a self-report measure designed specifically to assess deviant sexual fantasies among non-offending samples, to the MSI, a validated and widely used self-report measure of deviance among known sex offender populations.

It was hypothesized that there would be a significant positive relationship between self

reported levels of deviant sexual arousal and attentional bias toward sexually deviant words, measured by each participant's response latency for deviant words. Further, it was hypothesized that levels of self-reported deviant sexual arousal could be predicted by latencies to color name sexually deviant words (as compared to color naming non-deviant sexual words and neutral words).

CHAPTER 5: Methodology

Participants

The sample utilized in the current study included undergraduate students who were at least 18 years of age. The mounting evidence suggesting that deviant sexual interests are prevalent among non-offending samples (Hunt, 1974; Ogas & Gaddam, 2013), and university students (Maile & Jeglic, 2010) lends support for the use of this sample. Furthermore, the use of an undergraduate, nonoffending sample is also advantageous as a preliminary validation sample as they are not especially motivated to deny or conceal their sexual interests.

An a priori power analysis revealed that a sample size of at least 107 participants would provide sufficient statistical power to detect a medium to large effect size in this study, and at least 700 to detect a small effect. The final participant sample included 793 students who were enrolled in a Psychology course participating in the Research Experience Program. All Psychology students were invited to participate in the proposed study as an opportunity to fulfill their course research requirements, and it was advertised under the title, *An Examination of Interference Effects on a Color Naming Task.*

Overall, 570 females (72%) and 223 males (28%) participated in the study. Participant age ranged from 18 to 52 (M =20.5, SD = 3.91) with the majority of participants (87%) being between 18 and 26 at the time of their participation in the study. The majority (42%) of the sample was Hispanic (n = 337), with the remainder of the sample identifying as Black (18%; n=141), Caucasian (16%; n=124), Asian/Pacific Islander (15%; n=117), Middle Eastern (2%; n=15), American Indian/Alaskan Native (1%; n=7), and other (7%; n=57). Ninety-five percent of the sample was single and never been married, while 3.5% were currently married, 1% were divorced, and .4% were separated. The majority (41%) of the sample was comprised of college freshmen, while 28% were sophomores, 22% were juniors, and 9% were seniors.

Table 1

Participant Demographics

Demographic Variable	n	% of sample
Gender		
Male	223	28.1
Female	570	71.9
Ethnicity		
White, Non-Hispanic	124	15.6
Black, Non-Hispanic	141	17.8
Hispanic/Latino	332	41.9
American Indian/Alaskan Native	7	0.9
Asian/Pacific Islander	117	14.8
Middle Eastern	15	1.9
Other	57	7.2
Marital Status		
Single, Never Married	754	754
Married	28	28
Separated	3	3
Divorced	7	7
Grade Level		
Freshman	329	41.5
Sophomore	222	28
Junior	173	21.8
Senior	68	8.6
Annual Family Income Level		
Under $10,000	167	21.4
$10,000-$30,000	255	32.7
$30,000-$50,000	154	19.8
$50,000-$70,000	86	11
$70,000-$90,000	60	7.7
Greater than $90,000	57	7.3

Materials

The Emotional Stroop Test. The Emotional Stroop (Stroop, 1935; Ray, 1979) is a task that detects attentional interference via the measurement of response latencies when participants are presented with lists of words printed in different colors of ink. In its original format, participants state the color of the ink while inhibiting the meaning of the word. The response latency required to make a color selection is measured and interpreted as the degree of interference, such that longer latencies are interpreted as greater distraction by the content of the word. The emotional Stroop has been widely used among patients with affective disturbances and other populations to illustrate emotional interference (Williams et al., 1996). Stroop effects consistently demonstrate longer latencies to color name words to subjects' own concerns and pathologies as compared to neutral words. This effect has been demonstrated with alcohol disorders, cannabis abuse and dependence, eating disorders, and sleep disorders (Channon et al., 1988, Fairburn et al., 1991; Field, 2005; Sharma et al., 2001; Spiegelhalder et al, 2010).

In the current study, a modified, computerized version of the Stroop task was developed using three categories of words: neutral, sexual non-deviant (normative sexual words), and sexual deviant words, to empirically investigate the ability of the Stroop measure to distinguishing among individuals with various levels of deviant sexual interests. Words were presented randomly on a desktop computer using Empirisoft DirectRT (Jarvis, 2004) software. Respondents used color-specific keys on the keyboard (corresponding to red and blue), and the color-identification response latencies for each trial were recorded.

Participants first completed a practice trial to become familiar with the program. Directions were presented on the screen at the beginning of the task, which instructed participants to use the keyboard to indicate the color of each word presented on the screen as quickly and accurately as possible. Each trial started with a blank white screen, followed by the

presentation of a centered "+" symbol on the screen, which was then replaced by a word. The word remained on the screen until the participant made a color selection on the keyboard. They completed ten practice trials, after which they were provided time to ask questions or repeat the practice trial if necessary. The practice trials were then followed by the 132 experimental trials. In these experimental trials, participants were presented with 11 neutral words, 11 sexual normative words, and 11 deviant sexual words, each presented four times and in a new random order for each participant. Consistent with methodology from the extant Stroop literature (Cha, Najmi, Park, Finn, & Nock, 2010) and recent applications of Stroop methodology to sex offender populations (Price & Hanson, 2007; Price et al., 2013; Smith & Waterman, 2004), interference scores for sexual stimuli were calculated by subtracting mean RT for the neutral words from the mean RT for sexual words. Given that even normative, nonoffending samples demonstrate interference effects for sexual material on the Stroop task, it may be that sexual material is especially salient for individuals irrespective of their deviant sexual interests and/or behaviors. Therefore, we felt it necessary to isolate the effects of *deviant* sexual interests. Therefore, to calculate interference effects for deviant stimuli (i.e., deviant arousal-specific attentional bias), mean RT for sexual normative words was subtracted from the mean RT for sexual deviant words.

Word Development. The newly developed deviant and non-deviant sexual word lists were derived from the Smith and Waterman (2004) and Price and Hanson (2007) word lists, as well as several words generated in the research laboratory by doctoral and masters' level graduate students thought to be related to sex and sexual deviance. The set of words were pilot tested on multiple occasions.

First, a preliminary set of proposed deviant sexual words generated by the lab and previous research articles were independently rated for their sexual deviance on a Likert-type scale of 1-7 from not at all deviant (1) to most deviant (7) by doctoral and masters level graduate students in a sex offender research lab, as well as a sample of community members who were blind to the purposes of the study (*n*=30). Subsequent to a panel discussion, in which raters reported that the order of the words greatly influenced their deviancy ratings, a second pilot test was administered four months later, in which both remaining and new laboratory members, as well as community members, independently rated the words using the same scale. Additionally, sexual words were rated for their sexuality on a scale from not at all sexual (1) to most sexual (7). Lastly, to ensure that the proposed neutral words were indeed neutral and not sexual in nature, the neutral list was also rated on a scale from not at all sexual (1) to most sexual (7).

Sexual deviant words with a mean deviancy rating of at least 5.0 (e.g., rape, molest) and sexual (normative) words with a mean sexual rating of at least 4.0 (e.g., penetrate, masturbate) were retained for inclusion in the current study. Words with average ratings of 3 or below with regard to their respective sexual theme were discarded. Lastly, any words on the proposed neutral list with an average sexual or deviance rating greater than 1.0 were discarded from the neutral list and not utilized in the current study. A total of 33 words were included in the Stroop task (11 neutral, 11 sexual, 11 deviant).

In a separate but related validation study, the neutral words were investigated for their affective valence in accordance with the Self-Assessment Manikin (SAM) paradigm (Bradley & Lang, 1990). The SAM provides participants with a bipolar scale with figures that depict emotional valence on a scale from pleasant or happy (1) to unpleasant or unhappy (7). The normative ratings for the neutral words included in the study contained means of 2.40~5.72 (*SD*s

= 1.62-2.51). See app See Appendix A for the final version of the word stimuli utilized in this study.

Sexual Fantasy Questionnaire Validation. The Sexual Fantasy Questionnaire (SFQ; Gray, Watt, Hassan, & MacCulloch, 2003) is a 93-item measure that measures frequency of various sexual fantasies. It is especially useful for the current validation study, as it is the only known self-reported deviant sexual interest measure that was designed for use among nonoffending populations. As mentioned previously, research conducted using this measure has shown that as many as 87% of university students endorse deviant sexual interests (Maile & Jeglic, 2009; Maile, Maile, & Jeglic, 2010). Recent meta-analyses demonstrating a high level of agreement between physiological measures of DSA and self-report lend further support for the use of a self-report tool as a comparison measure for this study (Chivers, Seto, Lalumiére, Laan, & Grimbos, 2010).

For the purposes of this study, we were primarily interested in deviant sexual interests rather than behaviors, as the commission of deviant behavior is not a sufficient indicator of deviant sexual arousal patterns, as discussed above. Conversely, there may be many individuals in the sample that have deviant sexual interests but have not acted upon these interests due to situational factors and norms. As such, only the 93 fantasy items (and not behavioral items) of the SFQ were administered to our sample as a corroborating measure for which to compare Stroop performance. The questionnaire required participants to rate their perceived interest in various sexual fantasies ranging from normative to deviant. Participants rated their sexual interests on a Likert scale ranging from 0 to 4, with a 0 indicating never having fantasized about the corresponding content and a 4 indicating that the subject fantasizes about the content very frequently. The items fell into the following categories: normative sexual interests, rape,

pedophilia, sadism, masochism, non-contact offense related (i.e., voyeurism, exhibitionism), and other paraphilia-related fantasies (i.e., necrophilia, frotteurism). Scores on the eighteen items on the SFQ that specifically assessed deviant sexual fantasies were used to calculate an overall self-report deviance score. This total included items from the rape, pedophilia, and "other" categories and excluded items from the normative, sadism, masochism, and non-contact related categories. See Appendix B for a list of the eighteen items utilized to calculate overall deviance total score.

As mentioned previously, most self-report DSA measures have been designed for use with sexual offending populations, and the SFQ is therefore especially advantageous, as it is the only measure specifically designed to detect self-reported DSA among normative nonoffending samples. However, there is limited available data regarding the validity of the SFQ, and therefore it was crucial to assess the concurrent validity of the SFQ. Given that the ultimate goal of validating Stroop methodology is to eventually employ the assessment with sex offender populations, SFQ performance was compared to that of the Multiphasic Sex Inventory, one of the most commonly used self-report measures of DSA designed for known sexual offender populations (MSI; Nichols & Molinder, 1984). Therefore, the relationship between MSI, SFQ, and Stroop performance was explored among a small subsample of 18 male subjects.

The MSI asks about respondents' deviant sexual activities and interests. Of particular value, this measure includes social desirability and lie scales to detect respondent attempts to present themselves in an especially positive light, as well as a validity scale to check for consistency. Given that the MSI was originally developed for use among known sex offenders, it contains several items and scales relating to offense behavior and remorse. As this study utilized a nonoffending sample, the offense behavior-related items were omitted and only the deviant fantasy items were examined among a small subsample of male participants.

One of the benefits of utilizing the MSI is that it contains validity scales to assess for social desirability and subject response patterns. The Social Sexual Desirability (SSD) scale assesses normal sexual interest and drive to detect individuals attempting to respond in a socially desirable way or to appear "sexually hyper-normal" (Nichols & Molinder, 1984). Those who attempt to present themselves in a positive light deny normal sexual drives and interests and tend to score very low on this scale. Normative data suggests that the mean SSD score for normal male subjects is 30.875 (SD=3.215). Scores between 20 and 23 on the SSD scale (three standard deviations below the mean) suggest the person may be denying interest in sex, and scores beyond four standard deviations suggest the person is presenting in an asexual range. Descriptive analyses of the total SSD score among the male subsample revealed a mean score of 25.94 (SD=4.88, Range 16-32). Three participants scored below 23.

Another informative validity scale is the Sexual Obsessions Scale (SO). Nichols and Molinder (1984) reported that although it was created to measure sexual obsessions, it is helpful in detecting malingering. Normal adult samples tend to score a 2 or below, while sex offenders who are honest about their high interest in sex tend to score between 3 and 9. Scores less than 2 suggest the individual denying interest in sex, while scores greater than 17 suggest a "faking bad" response set. Descriptive analyses revealed a mean score of 3.28 (SD=2.99, Range 0-9) among the subsample in this study. The majority of participants (72%; n =13) scored between 0-3 on the SO.

Descriptive statistics revealed that 72.2% of the male sub sample ($n = 13$) endorsed deviant sexual interests on the MSI. This is consistent with the majority of the full sample (73%) endorsing some form of deviant sexual fantasy on the SFQ, as well as previous prevalence studies (Maile & Jeglic, 2009; Spada, Jeglic, & Calkins, 2013). Of the 13 who endorsed deviant

sexual interests, 69% (*n*=9) endorsed exhibitionism-related sexual fantasies, 46% (*n*=6) endorsed

rape-related sexual fantasies, and 38% (*n*=5) endorsed pedophilic sexual fantasies.

Table 7

Prevalence Rates and Percentages for MSI Deviant Sexual Fantasy Themes

MSI Sexual Interest Category	Male Subsample
Deviant Sexual Interests Overall	
Count	13
%	72.20%
Sexual Deviance: Rape Fantasy	
Count	6
% of overall subsample	33.30%
% of those endorsing deviant sexual interests	46.00%
Sexual Deviance: Child Molestation Fantasy	
Count	5
% of overall subsample	27.80%
% of those endorsing deviant sexual interests	38.00%
Sexual Deviance: Exhibitionism Fantasy	
Count	9
% of overall subsample	50.00%
% of those endorsing deviant sexual interests	69.00%
Atypical Sexual Outlet: Bondage & Discipline	
Count	5
% of overall subsample	27.80%
Atypical Sexual Outlet: Fetishism	
Count	1
% of overall subsample	5.60%

     Sub scores were calculated for the three MSI deviant sexual fantasy scales (child molest,

rape, and exhibitionism), as well as an overall MSI deviance total score. Next, correlational

analyses were conducted comparing the MSI scale scores to the SFQ scores and Stroop reaction

times. The results revealed a strong positive correlation between MSI child molestation sub score

and sexual deviance as measured by the SQF self-report measure, $r = .760$, $p = .000$. Similarly,

MSI rape fantasy sub score was significantly correlated with the total SFQ deviance score, r = .854, p = .000. However, neither the MSI total score, nor the rape or child molest sub scores were significantly correlated with deviant Stroop word latencies. MSI deviance total score was also significantly correlated with SFQ total score, r = .897, p = .000. High correlations between MSI and SFQ scores suggest that the SFQ is indeed a reliable measure for this validation study.

Procedure

Participants were asked to complete a computerized emotional Stroop task, in which the latencies to color name deviant sexual, non-deviant sexual, and neutral words, were recorded. More specifically, after signing up to participate in the confidential study through a web-based survey system, participants reported to the research laboratory, where they were seated in front of a computer and instructed on how to use the keyboard. All participants were assessed for color blindness. First, they were asked if they had any known histories of color blindness, and were then asked to demonstrate by pointing to colored keys on the keyboard. The computerized tasks were presented on one of two Dell computers, and participant responses were made using labeled keys on a Dell keyboard and a mouse. The Stroop tasks were run using the DirectRT response time research software, and the subsequent questionnaires were run using SurveyMonkey, a web-based survey program.

After obtaining informed consent, participants first completed a set of Stroop practice trials to become familiar with the software and response keys. Next, they completed the DSA Stroop task, where they were shown a series of words (sexual deviant, sexual normative, neutral) which were randomized and counterbalanced for order. Each image was presented four times. A fixation cross was presented prior to each stimulus for 1 second. For each word trial, participants were asked to select the appropriate key corresponding to the color that each word was printed

in, while ignoring the meaning of the word. Each word was presented in either red or blue type. Though no time limit was set for responses, participant response times were recorded for each trial, as well as whether the selected response was correct or incorrect.

Upon completion of the computerized Stroop task, participants then completed the computerized demographic questionnaire, as well as a computerized version of the Sexual Interest Questionnaire (SIQ; Maile, Maile, & Jeglic, 2010), which asks about perceived sexual interests. A subsample of 18 male participants then also completed a second survey, the Multiphasic Sex Inventory (MSI; Nichols & Molinder, 1984) for added concurrent validity, as this is a commonly used self-report measure of DSA. Only heterosexual male participants were administered the MSI, as it has been normed only with males and the questions are written in a male perspective. Rather than modifying the measure items to reflect female and homosexual interests, the decision was made to retain the measure in its original format and administer the measure only to a subsample of males. All sexual deviance self-report measures were administered *after* completion of the computerized Stroop task, so as to avoid priming or biasing the participants to respond in a particular way after being asked about their sexual interests and fantasies.

Analysis

The data analyses followed the procedural standards set forth by recent research on the Stroop paradigm, which included elimination of problematic responses (e.g., incorrect answers, random responding, individual and group outliers; Price & Hanson, 2007, Smith & Waterman, 2004, Price et al., 2013). Means and standard deviations for the reaction times (RT) of the three word categories were calculated, and Emotional Stroop response latencies (bias scores) were calculated as follows: For the sexual normative words, the mean reaction time for the neutral

words was subtracted from the sexual normative word RT to get a sexual bias score. For the sexual deviant words, the mean RT for the sexual normative words was subtracted from the mean RT for the sexual deviant words. This resulted in three response latencies to be compared to self-reported levels of deviant sexual arousal. Positive latencies reflect participants who were slower to color-name the words with the deviant sexual content due to interference effects.

To remain consistent with current practices in Stroop research (Cha et al., 2010), only trials with correct responses were included in the analysis. RT and Stroop bias scores were explored for outliers. Trials with response latencies ± two standard deviations from the participant's mean response latency were eliminated. Additionally, participants ($n = 13$) for whom the mean response latency was ± 2 standard deviations from the overall mean were excluded from the analysis. Lastly, participants ($n = 7$) for whom the error rate was two standard deviations above the error rate for all participants ($M = 1.95$, $SD = 4.76$) were also excluded from the analysis. When compared across all participants, the response latencies for sexual-deviant ($M = 603.93$, $SD = 169.54$), sexual-normative ($M = 607.73$, $SD = 171.36$), and neutral ($M = 599.82$, $SD = 165.26$) words were significantly different ($ps < .005$, $ds = .095$-$.983$).

CHAPTER 6: Results

The first aim was to identify the rates of self-reported deviant sexual arousal among a non-offending, university sample. The results indicated that while 26% of the sample endorsed only normative sexual interests, a majority of the sample ($n = 575$, 73%) endorsed some form of deviant sexual fantasy. Out of a possible total of 18 deviant items, the average number of deviant responses endorsed was 1.49 (SD=2.20, Range 1-16). The reported deviant items were used to comprise the following exclusive categories: rape/pedophilic, sadistic, masochistic, non-contact offense-related (i.e., voyeurism, exhibitionism), or other paraphilia-related fantasy (i.e., bestiality, necrophilia, frotteurism).

Of the 575 subjects who endorsed deviant sexual fantasies, 227 (39%) endorsed rape/child molestation related fantasies. Twenty-five percent (n=143) endorsed sadistic fantasies but did not endorse any rape/child molestation related fantasies, while twenty percent (n=114) exclusively endorsed masochistic fantasies. Eight percent of the sample (n=80) exclusively endorsed non-contacted offense related fantasies (i.e., voyeuristic, exhibitionistic), and 21 (4%) endorsed other deviant sexual fantasies (i.e., bestiality, necrophilia, frotteurism).

Table 2

Prevalence Rates and % within Gender for SFQ Deviant Sexual Fantasy Themes

Sexual Interest Category	Male (n=220)	Female (n=564)	Total (n=784)
Rape/Pedophilia			
Count	74	150	224
% Within Gender	33.6%	26.6%	28.6%
Other Paraphilia			
Count	8	13	21
% Within Gender	3.6%	2.3%	2.7%
Sadistic			
Count	38	102	140
% Within Gender	17.3%	17.4%	17.9%
Masochistic			
Count	15.0	98	113
% Within Gender	6.8%	17.4%	14.4%
Non-Contact Paraphilia			
Count	35.0	45	80
% Within Gender	15.9%	8.0%	10.2%
Normative Only			
Count	50	156	206
% Within Gender	22.4%	27.7%	26.3%

The second and third aims were to assess the response latencies for color naming sexually deviant words relative to the time required to color name normative sexual words and neutral words and to correlate Stroop word latency with self-reported deviance in an attempt to examine the validity and feasibility of using a modified Stroop to assess DSA. Descriptive statistics revealed that the mean RT for the neutral, normative sexual, and deviant sexual words across participants were 599.82 ms ($SD = 165.26$), 607.73 ms ($SD = 171.36$), and 603.93 ms ($SD = 169.54$), respectively. When broken down by gender, the mean RTs for females were significantly slower than the male RTs for all three word types ($p = .000$).

Table 3

Mean Stroop Reaction Times by Word Type and Gender

Category	Mean	*SD*
Male		
Neutral	550.69	138.24
Sexual Normative	557.56	144.97
Sexual Deviant	552.91	143.4
Female		
Neutral	606.32	150.18
Sexual Normative	614.29	155.28
Sexual Deviant	609.12	152.06
Overall Sample		
Neutral	599.82	165.26
Sexual Normative	607.73	171.36
Sexual Deviant	603.93	169.54

To correlate self-report with Stoop performance, an overall sexual deviance total score was first calculated for each participant, based on their responses to the self –report measure. A total of 18 deviant items were used to compute this score. Items related to sadism and masochism were excluded from the deviance score, given that recent research has deemed these as normative sexual interests. Given that participant responses could reflect 0-4, the total possible deviance score was 72 for these 18 items. Out of a total possible score of 72, the mean self-reported deviance total score for the sample was 20.52 ($SD = 4.79$). When compared by gender, males scored significantly higher ($M = 21.50$, $SD = 6.34$) on the self-report DSA measure than females ($M = 20.14$, $SD = 3.96$) overall, $t(1, 790) = 3.642$, $p = .000$.

A power analysis indicated the sample size for the present study provided sufficient statistical power (.96 with $\alpha = .05$, one-tailed tests) to detect small to medium effects. Results from correlational analyses revealed a significant positive correlation between total self-report

DSA score and the Stroop response latency for sexual deviant words, $r = .121, p = .001$. There was also a small but significant negative correlation between the broad deviance score and the Stroop latency for sexual words, $r = -.112, p = .001$. When analyzed by gender, the positive correlation between deviant response latency and DSA self-report total score remained for males ($r = .253, p = .000$), but not for females ($r = .058, p = .165$). Thus, among male participants there was a small to moderate relationship between their self-reported DSA and their response latencies on deviant Stroop word trials.

Given that the words selected during the pilot phase as being the most deviant correspond mostly to rape and child molestation, we next broke down the total DSA self-report score into a pedophilic interest self-report score and a rape interest self-report score. Additional correlational analyses were then run to investigate the relationships between these self-report sub scores and the deviant response latencies. The pedophilic interest items ($r = .076, p = .035$) and the rape interest items $(r = .111, p = .002)$ both correlated significantly with deviant word latency. Additionally, the rape interest items were also negatively correlated with sexual word latency ($r = -.125, p = .000$). When broken down by gender, the relationships again held true for the males but not for the females. For males, the pedophilic interest items ($r = .142, p = .035$) and the rape interest items ($r = .221, p = .001$) were both significantly correlated with deviant word latency. For females, neither self-report sub score was significantly correlated with deviant word latency (pedophilic interests $r = .064, p = .132$; rape interests $r = .061, p = .149$).

Table 4

Correlation Matrix for Male Participants

		SFQ Total Deviance Score	Rape Subscore	Pedophilic Subscore	Sexual Stroop Latency	Deviant Stroop Latency	Neutral Stroop Latency
SFQ Total Deviance Score	Pearson R	1	$.914^{**}$	$.547^{**}$	$-.149^{*}$	$.253^{**}$	$-.028$
	Sig.		.000	.000	.026	.000	.673
	N	223	223	220	223	223	223
Rape Subscore	Pearson R	$.914^{**}$	1	$.259^{**}$	$-.122$	$.221^{**}$	$-.022$
	Sig.	.000		.000	.070	.001	.740
	N	223	223	220	223	223	223
Pedophilic Subscore	Pearson R.	$.547^{**}$	$.259^{**}$	1	$-.122$	$.142^{*}$	$-.014$
	Sig.	.000	.000		.072	.035	.834
	N	220	220	220	220	220	220
Sexual Stroop Latency	Pearson R	$-.149^{*}$	$-.122$	$-.122$	1	$-.551^{**}$	$.133^{*}$
	Sig.	.026	.070	.072		.000	.048
	N	223	223	220	223	223	223
Deviant Stroop Latency	Pearson R.	$.253^{**}$	$.221^{**}$	$.142^{*}$	$-.551^{**}$	1	$-.041$
	Sig.	.000	.001	.035	.000		.541
	N	223	223	220	223	223	223
Neutral Stroop Latency	Pearson R	$-.028$	$-.022$	$-.014$	$.133^{*}$	$-.041$	1
	Sig.	.673	.740	.834	.048	.541	
	N	223	223	220	223	223	223

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

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

Table 5

Correlation Matrix for Female Participants

		Total Deviance Score	Rape Sub Score	Pedophilic Sub Score	Sexual Stroop Latency	Deviant Stroop Latency	Neutral Stroop Latency
Total Deviance Score	Pearson R	1	.869**	.427**	-.092*	.058	-.085*
	Sig.		.000	.000	.028	.165	.042
	N	569	568	563	569	569	569
Rape Sub Score	Pearson R	.869**	1	.046	-.129**	.061	-.056
	Sig.	.000		.272	.002	.149	.185
	N	568	568	562	568	568	568
Pedophilic Sub Score	Pearson R	.427**	.046	1	-.021	.064	-.034
	Sig.	.000	.272		.612	.132	.422
	N	563	562	563	563	563	563
Sexual Stroop Latency	Pearson R	-.092*	-.129**	-.021	1	-.528**	.083*
	Sig.	.028	.002	.612		.000	.048
	N	569	568	563	570	570	570
Deviant Stroop Latency	Pearson R	.058	.061	.064	-.528**	1	-.108**
	Sig.	.165	.149	.132	.000		.010
	N	569	568	563	570	570	570
Neutral Stroop Latency	Pearson R	-.085*	-.056	-.034	.083*	-.108**	1
	Sig.	.042	.185	.422	.048	.010	
	N	569	568	563	570	570	570

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

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

Given that there was a significant relationship between self-reported deviance and deviant word Stroop latencies and that this relationship held true among male subjects, the next step was to determine whether deviant Stroop latencies could predict self-reported deviance

scores. To investigate this, a linear regression was conducted, excluding the females from the model.

First, the data was evaluated in terms of whether or not it met the assumptions of the linear regression analysis. The assumptions of linearity, normality, and homoscedasticity were met. Next, the Durbin-Watson statistic (2.025) for the data was examined. Given that the value is so close to two, it was deemed an acceptable value. It demonstrated that there is independence of errors and no correlation between residuals. Thus, the independence of errors assumption was met.

As noted above, group outliers with significant errors and reaction times greater or less than two standard deviations from the mean were excluded. Upon examining case wise diagnostics, it was revealed that there were still six individuals remaining with standardized residuals greater than three standard deviations.

As others have noted, it is common for difficulties related to the assumptions of parametric tests to arise when working with RT data (Gress & Laws, 2009; Price et al., 2013). Further, Price and colleagues (2013) note that extreme values can provide valuable information about the research question being investigated, especially with regard to deviant sexual arousal. That is, we expect those with deviant sexual interests to be slower to color name on deviant words given the attentional bias, and it may therefore be problematic to treat those extreme scores as outliers. Given that this outlier was neither a data entry nor a measurement error (computer malfunction) as these had already been excluded, they were deemed likely to represent accurate data points and were included in the regression analysis.

The regression model was statistically significant, indicating that Stroop deviant word latency could significantly predict self-reported deviant sexual interest score, $F(1,221)= 15.128$, $p =.000$. Specifically, deviant Stroop word latency accounted for 6.4% of the variance in self-reported deviant sexual interests. The regression equation was: predicted self-reported deviance = 27.791 + 0.062 x (deviant Stroop word latency).

Table 6

Linear Regression Analysis for Predicting Self-Reported Deviance from Deviant Word Latency

Variable	B	SE B	β
Deviant Word Latency	0.062	0.16	0.253
R2		0.064	
F		15.128**	

**p < .01

CHAPTER 7: Discussion

This study serves as a preliminary but necessary step in examining the utility of Stroop attentional interference as an indicator of deviant sexual arousal. The study examined the prevalence of deviant sexual interests and the relationship between self-reported deviant sexual interests and Stroop latencies to neutral, sexual normative, and sexual deviant words among a non-offending, university undergraduate sample. The primary aim of the study was to determine whether deviant latencies could predict self-reported deviance, and more broadly, to examine the validity of the Stroop as a potential measure of DSA.

The results indicated that the majority (73%) of the undergraduate participants endorsed some type of deviant sexual interest, which is consistent with previous research examining the prevalence of deviant sexual interests among non-offending, college samples, and lends further support for the appropriateness of using this non-offending sample for this validation study (Maile & Jeglic, 2009; Spada, Jeglic, & Calkins, 2013). The findings highlight the difficulties inherent in defining sexual "deviance," as what is deemed deviant is socially constructed and varies over time. Additionally, they suggest that deviant sexual interests are not limited to sex offenders but that they are present among those in the general population. Further, it is likely that deviant sexual arousal itself exists on a continuum rather than being categorically present or absent.

Indeed, the *Diagnostic and Statistical Manual of Mental Disorders*, in versions as early as the *DSM-III*, acknowledges that there is a continuum between sexual health and sexual deviance and that the paraphilias, and all mental disorders, are not necessarily distinct entities (APA, 1980). The use of categorical classifications is most appropriate for situations in which all members of the category are homogenous and there are distinct boundaries between the

categories; however, no such clear boundaries exist for many mental disorders (including paraphilias), and group members are heterogeneous (APA, 2000). The high rate of deviant sexual interests among a general, non-offending population highlights the fact that these deviant fantasies and corresponding behaviors can be part of the sexual life of a normal, otherwise healthy individual and is not necessarily related to sexually abusive behavior (De Block & Adriaens, 2013).

For example, De Block and Adriaens (2013) note that "deviant" interests such as sexual coercion and masochistic fantasies of being beaten, raped, or humiliated are sexually arousing to some individuals. In fact, given the somewhat normative quality of some sexual interests thought to be deviant, the *DSM-IV* specifies that the fantasies, behaviors, or objects meet criteria for a paraphilia only when they lead to significant distress or impairment and are therefore different from "nonpathological use of sexual fantasies, behaviors, or objects as a stimulus for sexual excitement" (APA, 2000, pp.568). In the most recent revision of the DSM (*DSM-5*), the authors noted the individuals engaging in the atypical sexual fantasies or behaviors must either feel personal distress or cause distress or injury to others in order to be labeled as having a paraphilia (APA, 2013). To further emphasize the difference between having a deviant or atypical sexual interest and having an actual disorder, the *DSM-5* added the term disorder to describe the paraphilias (e.g., changed Sexual Masochism to Sexual Masochism Disorder). The fact that the DSM now stipulates that individuals may engage in atypical or deviant sexual interests without having a disorder lends support for the notion that deviant interests exist among non-offenders and may best be described as existing on a continuum. Additionally, it calls into question our definitions of exactly what qualifies as a "deviant" sexual interest. This makes identifying and measuring DSA particularly difficult, and yet increasingly necessary, as DSA is one of the

strongest predictors of sexual recidivism among sex offenders. Further, DSA may be best conceptualized as existing upon a continuum rather than as a categorical entity. For all these reasons, research should continue to refine measurement of DSA, and assessing DSA should be just one part of a multimodal, three-tiered risk assessment to include other dynamic factors and criminogenic needs, as well as the more historical, static factors (Hanson & Harris, 2007; Hanson & Thornton, 1999; Hanson et al., 2007).

The next aim was to utilize the Emotional Stroop Task for Deviant Sexual Arousal to assess response latencies for color naming sexually deviant words relative to non-deviant sexual words and neutral words. The results indicated that response latencies for sexual-deviant, sexual-normative and neutral words were significantly different, with the longest latencies overall for sexual-normative words, followed by sexual-deviant words and neutral words. In previous studies utilizing the Stroop to examine DSA, researchers compared performance between offender groups and no overall means were presented. That *all* individuals in the present study responded significantly slower for the sexual and deviant words as compared to the neutral words suggests that there may be something inherently unique about words with sexual content, such that they draw attention regardless of presence or degree of deviant sexual interests. Alternatively, it may be that the choice of words for assessing sexual deviance remains suboptimal. Future studies should continue to revise and improve word lists to derive the words that most accurately tap into DSA.

The third aim was to evaluate the validity of the Stroop as a measure of deviant sexual arousal by correlating performance on the Stroop task with the self-report measure of deviant sexual arousal. It was hypothesized that there would be a significant positive relationship between self-reported levels of deviant sexual arousal and attentional bias toward sexually

deviant words, measured by each participant's response latency for deviant words. The results indicated that there was indeed a significant relationship between deviant word latency and self-reported DSA. As self-reported deviant sexual interests increased, bias increased, as performance was slower for the deviant Stroop words. This is the first correlational Stroop study examining DSA using solely continuous variables. In the three other DSA Stroop studies to date, the researchers compared differences in Stroop performance among groups (e.g., sex offender versus non-offender). In the general Stroop literature however, some correlational work has been conducted. For instance, correlational studies of anxiety and Stroop performance demonstrate that self-reported anxiety as measured by state and trait anxiety questionnaires demonstrated that state anxiety was the better predictor of Stroop performance on anxiety-related words (Williams et al., 1996).

Of note, however, the relationship only held true for male participants and not for female participants. This is not especially surprising, as female sex offenders are exceedingly rare, and the sex offenses that *are* perpetrated by females are even less likely to be reported than for males, due in large part to rape myths (i.e., males cannot be raped; Kassing, Beesley, & Frey, 2005). Therefore, we know very little about the motivations behind sex offenses perpetrated by females and the characteristics of female sex offenders, in particular, whether or not deviant sexual arousal applies in the same manner as it does for males. Recent meta-analyses demonstrate that 62% of female sex offenders have been victims of sexual abuse themselves, and that sexual abuse lasted longer and was deemed more serious/violent than the abuse histories of females who did not become sex offenders (Colson, Boyer, Baumstarck, & Loundou, 2013). Importantly, the prevalence of sexual abuse victimization among female sex offenders is significantly higher than that of male sex offenders. Further 50% of female sex offenders have a psychiatric disorder,

significantly higher than that of male sex offenders and the general population (Colson et al., 2013).

Research has consistently demonstrated the link between DSA and sexual offending for males; however, this link has not been unequivocally established for females. Instead, explanations for female sexual offending have tended to focus on a dysfunctional manner of gratifying a need for emotional attachment (Duncan, 2010). Unlike that of male sex offender typologies, proposed female sex offender typologies (i.e., teacher/lover, male coerced/male accompanied) focus on an interpersonal connection with the victim rather than deviant sexual arousal (Matthews et al., 1989; Robertiello & Terry, 2007). For example, the teacher/lover typology describes females who abuse a victim, usually an adolescent, using their position of authority. Research states that teacher/lover offenders often possess cognitive distortions regarding the abuse, conceptualizing the sexual contact as one of kindness and love and not viewing themselves as a criminal, but rather externalizing the blame to others (Beech, Parrett, Ward, & Fisher, 2014; Robertiello & Terry, 2007). Nonetheless, the literature *does* also discuss what is referred to as the predisposed female sex offender, which describes a female sex offender who has been sexually or physically abused herself and may in fact be motivated by deviant and/or violent sexual fantasies (Matthews et al., 1989). However, it is unclear how many female sex offenders actually fit this description, as the development of typologies for female sex offenders have been exploratory in nature and based on very small sample sizes. More recent research has suggested that the most common typology is that of the teacher/lover, and that predisposed offenders may be much less common (Vandiver & Kercher, 2004). Given that the pathways to female sexual offending are currently not well known and there is little research to support a link between DSA and sexual recidivism among females, there is a also lack of

validated risk factors and risk assessment tools for females, including those that assess DSA (Gannon & Cortoni, 2010). Future research should continue to explore whether patterns of deviant sexual arousal are relevant risk factors for female sexual offenders.

Interestingly, the results of the current study also revealed a significant negative correlation between self-reported deviance as measured by the SFQ and sexual word latency, suggesting that as levels of self-reported deviance increased, responses were quicker for sexual words. Perhaps those with higher levels of sexual deviance find the purely sexual words to be less salient and the "shock value" of the sexual words is more pronounced in those with lower levels of deviant interests. Though it was hypothesized in this study that sexual words would be salient to all individuals, and therefore attempts were made to isolate the effects of the deviant content from that of the sexual content, perhaps greater levels of deviance result in less attentional pull from purely sexual normative words. There is evidence to suggest that those with deviant sexual interests who meet the threshold for a paraphilia may become incapable of arousal except when the subject of their paraphilia is present (APA, 2000). There is also evidence to suggest that those with fetishism also require the preferred object in order to experience sexual excitement (APA, 2000). Therefore, it is possible that this may extend to more implicit attentional bias. Future studies might replicate this study with highly deviant sex offender populations to examine whether their performance on purely sexual words is faster (i.e., less attentional bias) as their levels of deviance increase.

The final aim was to determine whether response latencies for deviant words on the Stroop task could significantly predict self-reported levels of deviant sexual arousal using regression analyses. Regression analyses indicated that deviant word response latencies could significantly predict male self-reported deviant sexual interests. Previous studies have

investigated the Stroop as a measure of DSA by making group comparisons (i.e., sex offender, non-offender; Price et al., 2013; Price & Hanson, 2007; Smith & Waterman, 2004) to determine whether there are significant differences in interference scores, with mixed results. It can be argued, however, that the groups were not homogenous with regard to the presence or degree of deviance. That is, when comparing offenders to non-offenders, it is likely that both groups are comprised of individuals with and without deviant sexual interests. Therefore, the ability to detect significant differences in Stroop performance is limited, as group membership may not be the most accurate concurrent validation measure. Additionally, the dichotomization of variables may further limit the power to detect significance.

In summation, the results of the current study provide support for the Stroop as a valid and feasible tool for assessing deviant sexual arousal; however, the small effect size suggests that there are other factors influencing DSA. This is consistent with the extant literature on the role of deviant sexual fantasies in sexual offending and sexual recidivism, which suggests that assessing DSA is an extremely aspect of risk assessment, but that it is also interrelated with various other important static and dynamic factors (Harris, 2014; Harris & Hanson, 2010; Bartels & Gannon, 2011). In fact, as mentioned previously, current best practices in sex offender risk assessment includes the consideration of 13 stable dynamic factors, seven acute dynamic factors, and 10 static factors (Harris, 2014). Furthermore, a variety of other motivational and situational (i.e., victim access) factors may contribute to the relationship between DSA and sexual offending (Finkelhor, 2009).

The current study utilized self-reported deviance, as opposed to group membership or offense behavior, as a measure of concurrent validity, and the results suggest the Stroop is in fact a viable tool for assessing DSA, as Stroop performance could significantly predict self-reported

DSA among male participants. These results are promising, as it seems crucial that the basic methodology of the Stroop be accurate in detecting DSA before attempts can be made with any degree of confidence to compare offender groups. This sample was not motivated to lie about their sexual interests and did indeed endorse a great degree of deviant interests. The results suggest that if an individual consciously and effortfully reports that they possess deviant interests, an automatic, implicit measure, the Stroop, can corroborate that self-report. The next step would therefore be to determine 1) whether the Stroop can detect DSA when an individual *is* attempting to lie or conceal it (i.e., offender samples) and 2) whether the Stroop performance has predictive utility with regard to sexual recidivism.

Clinicians are routinely asked to evaluate sexual recidivism risk, and evaluation of deviant sexual arousal is one critical component of that evaluation. In fact, high stakes decisions regarding offenders' freedom including but not limited to decisions about treatment, supervision, and Sexually Violent Predator (SVP) civil commitment are based upon evaluations of deviant sexual arousal (Perillo, Spada, Calkins, & Jeglic, 2014). There is some concern that paraphilia diagnoses (indicating the presence of pathological sexual deviance) may be applied too broadly when there is evidence simply of sexual offense behavior (Perillo et al., 2014). However, as mentioned previously, deviant sexual behavior is *not* sufficient evidence of deviant sexual arousal (First & Halon, 2008).

What's more, these consequences of these decisions are extremely costly, both financially and otherwise. For example, recent estimates suggest that civil commitment of one offender for a year costs approximately $100,000 (Rappleye, 2012). Given the possible consequences of sex offender risk assessment, to both the offender and society, the most accurate assessment of those sex offenders at highest risk to commit future acts of sexual violence is

critical. This study may serve as an initial step in working toward perfecting an assessment tool that can be used among clinicians to evaluate deviant sexual arousal.

The current gold standard for assessing DSA, the penile plethysmograph (PPG), while well-established in the literature, is expensive and has recently come under scrutiny for ethical and methodological complains, as well as questions about its validity and ability to be distorted or faked. Self-report assessments, while much cheaper and considerably less invasive, are even more vulnerable to distortion than the PPG. Therefore, it has become increasingly necessary to develop an alternative method of assessing DSA and recently, cognitive measures have been proposed. The development of a system of standardized cognitive tasks would be an easy and cost-efficient addition, or even possibly a substitute for the costly, invasive, and labor-intensive PPG (O' Ciardha & Gormley, 2009). The Stroop measure is economical, easy to administer, and non-invasive. The current study provides support for its use in detecting DSA and suggests that, with future improvements, it has the potential to provide a faster and more ethically sound alternative for evaluating DSA.

CHAPTER 8: Limitations and Directions for Future Research

There were several methodological limitations to the present study. First we utilized self-report measures as the corroborating measure to which Stroop performance was compared. Although research has demonstrated high levels of agreement between physiological measures of DSA and self-report (Chivers et al., 2010), there remains the possibility of measurement error when using this type of assessment. For one, it assumes that the respondents are insightful and truthful about their own sexual interests which, for some, may not be the case. Despite the anonymity of their participation, it is still possible that participants may respond in a socially desirable pattern. Those respondents who may possess DSA but either conceal it or are otherwise unaware of it were categorized for our purposes as non-deviant and their self-report was then compared to their Stroop performance.

This is problematic and could potentially be improved by adding additional measures of DSA that are less susceptible to social desirability and awareness. It is worth noting that some of the social desirability limitations were overcome by supplementing the SFQ by administering the MSI and its social desirability subscales, which suggested that our subsample of males were not particularly motivated to deny their interests in sex or attempting to present a nonsexual image. Despite issues with the utilization of the plethysmograph as a measure of DSA, it still remains the gold standard in the field. Thus, future research should include regression analyses comparing Stroop performance to physiological measures like the penile plethysmography.

The sample utilized in this study was a non-offending, university student sample. A majority of the sample endorsed deviant sexual interests, making them an appropriate sample for the study. However, they are not representative of the general population, nor are they representative of sex offenders specifically. Future research validating the Stroop as a measure of

DSA should replicate this study with larger, more representative sample in terms of age and background. Future research should also include application of this Stroop paradigm to sex offender samples.

As mentioned previously, perhaps the most critical quality of a measure of deviant sexual arousal is its ability to predict future sexual recidivism, as the ultimate goal is to use this measure as part of a multimodal risk assessment. Given that this is a non offending sample we have no reason to believe that these individuals would offend sexually in the future and thus we cannot assess predictive validity. Therefore, future research should also aim to empirically examine the predictive accuracy of Stroop performance in predicting sexual recidivism by including samples of repeat sex offenders with first time offenders.

Appendix A

Modified Emotional Stroop Word Stimuli by Category

Neutral	Sexual Normative	Sexual Deviant
Coffee	Arousal	Abuse
Color	Erotic	Choke
Cup	Foreplay	Cunt
Leaves	Lust	Humiliate
Pizza	Masturbate	Incest
Soda	Naked	Molest
Speak	Orgasm	Rape
Television	Passion	Stalk
Toothbrush	Penetrate	Torture
Train	Pornography	Violate
Windy	Sex	Whip

Appendix B

Sexual Fantasy Questionnaire (SFQ) Items Comprising Total Deviance Score

Item Description
1. Pre-puberty boys
2. Pre-puberty girls
3. Adolescent boys
4. Adolescent girls
5. Physically hurting the person you are having sex with
6. Physically injuring the person you are having sex with
7. Strangling or asphyxiating others without killing
8. Torturing others
9. Raping somebody
10. Humiliating somebody
11. Drugging or sedating another person for sexual reasons
12. Physically attacking someone
13. Forcing somebody to do something non-sexual against their will
14. Forcing somebody to have sex against their will
15. Sex while threatening someone with a weapon
16. Stalking or secretly following somebody
17. Sex with a dead person
18. Frottage (rubbing yourself against others)

*For each item, participants were given the instructions, "Please report your level of sexual <u>interest</u> in relation to the following things or situations." They indicated their interest on the following Likert scale from 0-4.

0 = Never fantasize about
1 = Have fantasized about once
2 = Have fantasized about several times
3 = Fantasize about frequently
4 = Fantasize about very frequently

References

Abel, G. G., Rouleau, J. L., Lawry, S. S., Barrett, D. H., & Camp, N. L. (1990). *A non-invasive, physiological measure of adult sexual preference*. Paper presented at the annual meeting of the International Academy of Sex Research, Sigtuna, Sweden, Aug.

Abel, G. G., Huffman, J., Warberg, B., & Holland, C. L. (1998). Visual reaction time and plethysmography as measures of sexual interest in child molesters. *Sexual Abuse: A Journal of Research and Treatment, 10*, 81-95. doi:10.1177/107906329801000202

Abel, G. G., Jordan, A., Rouleau, J. L., Emerick, R., Barboza-Whitehead, S., & Osborn, C. (2004). Use of visual reaction time to assess male adolescents who molest children. Sexual Abuse: A Journal of Research and Treatment, 16, 25-265. doi:10.1177/107906320401600306

Akerman, G. (2008). The development of a fantasy modification programme for a prison-based therapeutic community. *International Journal of Therapeutic Communities, 29*, 180-188.

Akerman, G. & Beech, A. R. (2012). A systematic review of measures of deviant sexual interest and arousal. *Psychiatry, Psychology, and Law, 19*, 118-143. doi:10.1080/13218719.2010.547161

American Psychological Association. (1999). *Standards for Educational and Psychological Testing* (2nd ed). American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed). American Psychiatric Publishing.

Andrews, D. A. & Bonta, J. (1994). *The psychology of criminal conduct.* Cincinnati, OH: Anderson.

Babchisin, K. M., Nunes, K. L., & Hermann, C. A. (2013). The validity of implicit association

test (IAT) measures of sexual attraction to children: A meta-analysis. *Archives of Sexual Behavior, 42*, 487-499. doi:10.1007/s10508-012-0022-8

Barbaree, H.E., & Marshall, W.L. (1988). Deviant sexual arousal, demographic and offense history variables as predictors of reoffense among child molesters and incest offenders. *Behavioral Sciences & the Law, 6*, 267-280

Barbaree, H. E. & Marshall, W. L. (1989). Erectile responses among heterosexual child molesters, father-daughter incest offenders, and matched non-offenders: Five distinct age preference profiles. *Canadian Journal of Behavioral Science, 21*, 70-82. doi:10.1037/h0079791

Barker, J. G. & Howell, R. J. (1992). The plethysmograph: A review of recent literature. *Bulletin-American Academy of Psychiatry & Law, 20*, 13-25.

Bartels, R. M. & Gannon, T. A. (2011). Understanding the sexual fantasies of sex offenders and their correlates. *Aggression and Violent Behavior, 16*, 551-561. doi:10.1016/j.avb.2011.08.002

Baumgartner, J. V., Scalora, M. J., & Huss, M. T. (2002). Assessment of the Wilson Sex Fantasy Questionnaire among child molesters and nonsexual forensic offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*, 19-30. doi:10.1177/107906320201400102

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

Beech, A. R., Kalmus, E., Tipper, S., Baudouin, J., Flak, V., & Humphreys, G. W. (2008). Children induce an enhanced attentional blink in child molesters. *Psychological Assessment, 20*, 397-402. doi:10.1037/a0013587

Beech, A. R., Parrett, N., Ward, T., & Fisher, D. (2014). Assessing female sex offenders'
motivations and cognitions: An exploratory study. *Psychology, Crime, & Law, 15*, 201-
216. doi:10.1080/10683160802190921

Best, J. (2011). Constructing deviance. In C. D. Bryant (Ed.), *The routledge handbook of deviant
behavior* (pp. 17-23). New York, NY: Routledge/Taylor & Francis Group.

Blasko, B. L., Jeglic, E. L., & Mercado, C. C. (2010). Are actuarial risk data used to make
determinations of sex offender risk classification?: An examination of sex offenders
selected for enhanced registration and notification. *International Journal of Offender
Therapy and Comparative Criminology, 55*, 676-692. doi: 10.1177/0306624X10372784

Boccaccini, M. T., Murie, D. C., Caperton, J. D., & Hawes, S. W. (2009). Field validity of the
Static-99 and MnSOST-R among sex offenders evaluated for civil commitment as
sexually violent predators. *Psychology, Public Policy, and Law, 14*, 278-314.
doi:10.1037/a0017232

Boccaccini, M. T., Murie, D. C., Mercado, C. C., Quesada, S., Hawes, S., Rice, A. K., & Jeglic,
E. L. (2012). *Implications of Static-99 field reliability findings for score use and
reporting. Criminal Justice & Behavior, 39*, 42-57. doi: 10.1177/0093854811427131

Cha, Najmi, Park, Finn, & Nock, 2010

Channon, S., Hemsley, D., & de Silva, P. (1988). Selective processing of food words in anorexia
nervosa. *Journal of Clinical Psychology, 27*, 259-260.
doi: 10.1111/j.2044-8260.1988.tb00782.x

Chivers, M. L., Seto, M. C., Lalumiére, L., & Grimbos. (2010). Agreement of self-reported and
genital measures of sexual arousal in men and women: A meta-analysis. *Archives of
Sexual Behavior, 39*, 5-57. doi: 10.1007/s10508-009-9556-9

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 1983. doi:10.1177/014662168300700301

Cohen, M. & Jeglic, E. L. (2007). Sex offender legislation in the United States: What do we know? *International Journal of Offender Therapy and Comparative Criminology, 51*, 369-383. doi: 10.1177/0306624X06296235

Colson, M., Boyer, L., Baumstarck, K., & Loundou, A. D. (2013). Female sex offenders: A challenge to certain paradigms meta-analysis. *Sexologies, 22*, e109-e117. doi:10.1016/j.sexol.2013.05.002

Crooks, V. L., Rostill-Brookes, H., Beech, A. R., & Bickley, J. A. (2009). Applying rapid serial visual presentation to adolescent sexual offenders: Attentional bias as a measure of deviant sexual interest? Sexual Abuse: *A Journal of Research and Treatment, 21*, 135-148. doi:10.1177/1079063208328677

De Block, A. & Adriaens, P. R. (2013). Pathologizing sexual deviance: A history. *Journal of Sex Research, 50*, 276-298. doi:10.1080/00224499.2012.738259

Dubner, A. E. & Motta, R. W. (1999). Sexually and physically abused foster care children and posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 67*, 367-373.

Duncan, K. A. (2010). *Female Sexual Predators: Understanding Them to Protect our Children and Youths*. California: Praeger.

Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse: A Journal of Research and Treatment, 24*, 5-28. doi:10.1177/1079063211403164

Epperson, D. L., Kaul, J. D., Huot, S., Goldman, R., & Alexander, W. (2003). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) Technical Paper: Development, Validation, and Recommended Risk Level Cut Scores.* Retrieved August 7, 2012 from http://www.psychology.iastate.edu/~./!old%20accounts/dle/mnsost_download.htm.

Fairburn, C. G., Cooper, P. J., Cooper, M. J., McKenna, F. P., & Anastasiades, P. (1991). Selective information processing in bulimia nervosa. *International Journal of Eating Disorders, 10*, 415-422. doi:10.1002/1098-108X(199107)10:4<415::AID-EAT2260100406>3.0.CO;2-R

Farren, L., & Barnett, G. (in prep). *Validation of a psychometric measure of sexual interests, the 'my private interest scale' (mpi), with a sample of convicted English and Welsh male sexual offenders*.

Field, M. (2005). Cannabis 'dependence' and attentional bias for cannabis-related words. *Pharmacology, 16*, 473-476. doi: 10.1097/00008877-200509000-00021

Finkelhor, D. (2009). The prevention of childhood sexual abuse. *The Future of Children, 19*, 169-194.

First, M. B. & Halon, R. L. (2008). Use of DSM paraphilia diagnoses in sexually violent predator commitment cases. *The Journal of the American Academy of Psychiatry and the Law, 36*, 443-454.

Freund, K. (1977). Psychophysiological assessment of change in erotic preferences. *Behavioral Research and Therapy*, *15*, 297-301.
doi: http://dx.doi.org/10.1016/0005-7967(77)90025-0

Freund, K. & Watson, R. J. (1991). Assessment of the sensitivity and specificity of a

phallometric test: An update of phallometric diagnosis of pedophilia. *Psychological Assessment, 3*, 254-260.

Furby, L., Weinrott, M. R., & Blackshaw, L. (1989). Sex offender recidivism: A review. *Psychological Bulletin, 105*, 3-30.

Gallagher, C.A., Wilson, D.B., Hirschfield, P., Coggeshall, M.B., & MacKenzie, D.L. (1999). A quantitative review of the effects of sex offender treatment on sexual reoffending. *Corrections Management Quarterly, 3*, 19-29.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*, 575-607.

Giotakos, O. (2006). A combination of viewing reaction time and incidental learning task in child molesters, rapists, and control males and females. *Annals of General Psychiatry, 5*, S114. doi:10.1186/1744-859X-5-S1-S114

Gosselin, C. & Wilson, G. (1980). *Sexual variations*. London: Faber and Faber.

Greenberg D. M., Firestone P., Bradford J. M., & Curry S. (2000). Recidivism of child molesters: A study of victim relationship with perpetrator. *Child Abuse & Neglect, 24*, 1485-1494. doi:10.1016/S0145-2134(00)00197-610.1016/S0145-2134(00)00197-6

Gress, C. L. Z., Anderson, J. O. & Laws, D. R. (2013). Delays in attentional processing when viewing sexual imagery: The development and comparison of two measures. *Legal and Criminal Psychology, 18*, 68-82. doi: 10.1111/j.2044-8333.2011.02032.x

Gress, C. L. Z. & Laws, D. R. (2009). Cognitive modeling of sexual arousal and interest: Choice reaction time measures. In D. Thornton & D. R. Laws (Eds.), Cognitive approaches to the assessment of sexual interest in sexual offenders (pp. 177-201). Chichester, UK: Wiley.

Hall, G. C. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology, 63*, 802-809.

Hanson, R. K. & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362. doi:10.1037/0022-006X.66.2.348

Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*, 169-194. doi:10.1177/107906320201400207

Hanson, R. K. & Harris, A. J. R. (2000). Where should we intervene?: Dynamic predictors of sexual offense recidivism. *Criminal Justice & Behavior, 27*, 6-35. doi:10.1177/0093854800027001002

Hanson, R. K. & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 13*, 105-122. doi: 10.1177/107906320101300204

Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The dynamic supervision project*. Public Safety Canada. Retrieved from: http://www.static99.org/pdfdocs/hansonharrisscottandhelmus2007.pdf

Hanson, R. K. & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*, 1154-1163. doi:10.1037/0022-006X.73.6.1154

Hanson, R. K. & Morton-Bourgon, K. E. (2007). *The accuracy of recidivism risk assessments for*

sexual offenders: A meta-analysis. (Corrections User Report No. 2007-01). Ontario, Canada: Public Safety and Emergency Preparedness Canada.

Hanson, R. K., Scott, H., & Steffy, R. A. (1995). A comparison of child molesters and nonsexual criminals: Risk predictors and long-term recidivism. *Journal of Research in Crime and Delinquency, 32*, 325-337. doi: 10.1177/0022427895032003004

Hanson, R. K., & Thornton, D. (1999). STATIC-99: Improving actuarial risk assessments for sex offenders (User Report No. 1999-02). Ottawa: Department of the Solicitor General of Canada.

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised*. Toronto, Canada: Multi-Health Systems.

Harris, A. J. (2014, March). *Stable-2007 and Acute-2007 Training: The Basics of Dynamic Risk Assessment*. Full day training presented at Rutgers University Inn and Conference Center, New Brunswick, New Jersey.

Harris, A. J. R. & Hanson, R. K. (2010). Clinical, actuarial and dynamic risk assessment of sexual offenders: Why do things keep changing? *Journal of Sexual Aggression, 16*, 296-310. doi:10.1080/13552600.2010.494772

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiére, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-425. doi: 10.1037/1040-3590.15.3.413

Heil, P., Ahlmeyer, S., & Simons, D. (2003). Crossover sexual offenses. *Sexual Abuse, A Journal of Research and Treatment, 15*, 221-236.

Hewitt J. A. (2002) *The architecture of thought: A new look at human evolution.* Holmhurst House. Retrieved from sexandphilosophy.co.uk.

Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment, 7*, 9-24. doi: 10.1177/107906329500700104

Hua, C. (2013). Sex weekend examines sexual culture. *Yale Daily News*. Retrieved from http://yaledailynews.com/blog/2013/03/04/sex-weekend-examines-sexual-culture/

Hunt, M. (1974). Sexual behavior in the 70's. Chicago: Playboy Press.

Imhoff, R., Schmidt, A. F., Nordseik, U., Luzar, C., Young, A.W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior, 39*, 1275-1288. doi:10.1007/s10508-009-9595-2

Jarvis, B. (2004). DirectRT (Version 2004) [Computer software]. NewYork, NY: Empirisoft Corp.

Kalichman, S. C., Henderson, M. C., Shealy, L. S., & Dwyer, M. (1992). Psychometric properties of the multiphasic sex inventory in assessing sex offenders. *Criminal Justice and Behavior, 19*, 384-396. doi:10.1177/0093854892019004003

Kalichman, S. C., Syzmanowski, D., McKee, G., Taylor, J., & Craig, M. E. (1989). Cluster analytically derived MMPI profile sub-groups of incarcerated adult rapists. *Journal of Clinical Psychology, 45*, 149-155.

Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*, 193-217. doi:10.1016/j.avb.2003.12.002

Kassing, L. R., Beesley, D., & Frey, L.L. (2005). Gender role conflict, homophobia, age, and education as predictors of male rape myth acceptance. *Journal of Mental Health Counseling, 27,* 311-328.

Kleban, H., Chesin, M. S., Jeglic, E. L., & Mercado, C. C. (2012). An exploration of crossover

sexual offending. *Sexual Abuse: A Journal of Research and Treatment, 25*, 427-443. doi:10.1177/1079063212464397

Knight, R. A. (2008). *MIDSA clinical manual*. Bend, OR: Augur Enterprises. Retrieved 6 July 2012, from www.midsa.us.

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*, 150-175. doi: 10.1177/0093854894021001010

Laws, D. R. (2009). Penile plethysmography: Strengths, limitations, innovations. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp. 7-29). Chichester, UK: Wiley.

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

Leitenberg, H. & Henning, K. (1995). Sexual fantasy. *Psychological Bulletin, 3*, 469-496.

Levenson, J. S., Fortney, T., & Baker, J. N. (2010). Views of sexual abuse professionals about sex offender notification policies. *International Journal of Offender Therapy and Comparative Criminology, 54*, 150-168. doi: 10.1177/0306624X08327785

Maile, C. & Jeglic, E.L (2009, October). *Prevalence of paraphilic fantasy in a non-offending sample*. Poster session presented at the meeting of the Association for the Treatment of Sexual Abusers, Dallas, Texas.

Malcom, P. B., Andrews, D. A., & Quinsey, V. L. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence, 8*, 486-501.

Marshall, W. L. (2006). Clinical and research limitations in the use of phallometric testing with

sexual offenders. *Sexual Offender Treatment, 1*. Retrieved 10 August 2012 from http://www.sexual-offender-treatment.org/marshall.html.

Matthews, J. K., Mathews, R., & Speltz. K. (1991). Female sexual offenders: A typology. In M.Q. Patton (Ed.). *Family sexual abuse: Frontline research and evaluation*, Sage Publications, Newbury Park (1991).

McGrath, R., Cumming, G., Burchard, B., Zeoli, S., & Ellerby, L. (2010). *Current practices and trends in sexual abuser management: The safer society 2009 North American survey*. Brandon, Vermont: Safer Society Press.

Mokros, *A.*, Dombert, B., Osterheider, M., Zappala, A., & Santtila, P. (2010). Assessment of pedophilic sexual interest with an attentional choice reaction time task. *Arch Sex Behav, 39*, 1081-1090. doi:10.1007/s10508-009-9530-6

Moster, A., Wnuk, D. W., & Jeglic, E. L. (2008). Cognitive Behavioral Therapy interventions with sex offenders. *Journal of Correctional Health Care, 14*, 109-121. doi:10.1177/1078345807313874

Nichols, H. R. & Molinder, L. (1984). *The Multiphasic Sex Inventory Manual.*

Nichols, H. R. & Molinder, L. (1996). *Multiphasic Sexual Inventory II Handbook*. Tacoma, WA: Nichols & Molinder Associates.

Ó Ciardha, C. (2011). A theoretical framework for understanding deviant sexual interest and cognitive distortions as overlapping constructs contributing to sexual offending against children. *Aggression and Violent Behavior, 16*, 493-502. doi:10.1016/j.avb.2011.05.001

Ó Ciardha, C. & Gormley, M. (2012). Using a pictorial-modified Stroop task to explore the sexual interests of sexual offenders against children. *Sexual Abuse: A Journal of Research and Treatment, 24*, 175-197. doi: 10.1177/1079063211407079

O'Donohue, W. & Letourneau, E. (1992). The psychometric properties of the penile tumescence assessment of child molesters. *Journal of Psychopathology and Behavioral Assessment, 14*, 123-174.

Odeshoo, J. R. (2004). Of penology and perversity: The use of penile plethysmography on convicted child sex offenders. *Temple Political and Civil Rights Law Review 14*, 1.

Ogas, O. & Gaddam, S. (November, 2013). *A billion wicked thoughts: What the internet reveals about sexual desire*. Paper presented at the annual conference of the Association for the Treatment of Sexual Abusers, Chicago, IL. http://tinyurl.com/nxjrv5f

Olver, M. E., Nicholaichuk, T., Gu, D., & Wong, S. C. P. (2012). Sex offender treatment outcome, actuarial risk, and the aging sex offender in Canadian corrections: A long-term follow-up. *Sexual Abuse: A Journal of Research and Treatment, 25*, 396-422. doi:10.1177/1079063212464399

Olver, M. E., Wong, S. C., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the violence risk scale-sexual offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318, 329.

Osborn, C., Abel, G. G., & Warberg, B. W. (1995). *The Abel Assessment: Its comparison to plethysmography and resistance to falsification.* Paper presented at the annual conference of the Association for the Treatment of Sexual Abusers, New Orleans, LA, Oct.

Perillo, A. D., Spada, A. H., Calkins, C., & Jeglic, E. L. (2014). Examining the scope of questionable diagnostic reliability in sexually violent predator (SVP) evaluations. *International Journal of Law and Psychiatry, 2*, 190-197. doi:10.1016/j.ijlp.2013.11.005

Prentky, R. A. & Burgess, A. W. (2000). *Forensic Management of Sexual Offenders*. New York, NY: Kluwer Academic/Plenum Publishers.

Price, S. A. & Hanson, R. K. (2007). A modified Stroop task with sexual offenders: Replication

    of a study. *Journal of Sexual Aggression, 13*, 203-216. doi:10.1080/13552600701785505

Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism.

    Journal of Interpersonal Violence, 85, 85-105. doi: 10.1177/088626095010001006

Quinsey, V. L., Rice, M. E., Harris, G. T., & Reid, K. S. (1993). Conceptual and measurement

    issues in the phylogenetic and ontogenetic development of sexual age preference in

    males. In H. E. Barbaree , W. L. Marshall, & S. M. Hudson, (Eds.), *The juvenile sex*

    offender (pp. 143-163). New York: Guilford.

Rappleye, H. (2012). Releasing sex offenders. *The Crime Report* (2012, April 17). Retrieved

    from http://www.thecrimereport.org.

Rice, M. E., Harris, G. T., Lang, C., & Chaplin, T. C. (2008). Sexual preferences and recidivism

    of sex offenders with mental retardation. *Sexual Abuse: A Journal of Research and*

    Treatment, 20, 409-425. doi: 10.1177/1079063208324662

Robertiello, G. & Terry, K. J. (2007). Can we profile sex offenders? A review of sex offender

    typologies. *Aggression and Violent Behavior, 12*, 508-518.

    doi:10.1016/j.avb.2007.02.010

Sandler, J. C., Freeman, N. J., & Socia, K. M. (2008). Does a watched pot boil? A time-series

    analysis of New York state's sex offender registration and notification law. *Psychology,*

    Public Policy, & the Law, 14, 284-302. doi: 10.1037/a0013881

Schlank, A. M. (1995). The utility of the MMPI and the MSI for identifying a sexual offender

    typology. *Sexual Abuse: A Journal of Research and Treatment, 2*, 185-194.

    doi:10.1177/107906329500700302

Serin, R. C., Mailloux, D. L., & Malcom, P. B. (2001). Psychopathy, deviant sexual arousal, and

recidivism among sexual offenders. *Journal of Interpersonal Violence, 16*, 234-245. doi: 10.1177/088626001016003004

Serin, R. C., Malcom, B. P., Khanna, A., & Barbaree, H. E. (1994). Psychopathy and deviant sexual arousal in incarcerated sexual offenders. *Journal of Interpersonal Violence, 9*, 3-11.

Seto. M. C. (2001). The value of phallometry in the assessment of male sex offenders. *Journal of Forensic Psychology Practice, 1*, 65-75. doi: 10.1300/J158v01n02_05

Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC: American Psychological Association.

Seto, M. C. (2009). Pedophilia. *Annual Review of Clinical Psychology, 5*, 391-407. Doi: 10.1146/annurev.clinpsy.032408.153618

Seto, M. C., Lalumière, M. L., & Blanchard, R. (2000). The discriminant validity of a phallometric test for pedophilic interest among adolescent sex offenders against children. *Psychological Assessment, 12*, 319-327. doi: 10.1037/1040-3590.12.3.319

Seto, M. C., Lalumière, M. L., & Kuban, M. (1999). The sexual preferences of incest offenders. *Journal of Abnormal Psychology, 108*, 267-272.

Sharma, D., Albery, I. P., & Cook, C. (2001). Selective attentional bias to alcohol-related stimuli in problem drinkers and non-problem drinkers. *Addiction, 96*, 285-295. doi: 10.1080/09652140020021026

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

Smith, G. & Fischer, L. (1999). Assessment of juvenile sex offenders: Reliability and validity of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and*

Treatment, 11, 207-216. doi:10.1177/107906329901100304

Smith, P. & Waterman, M. (2004). Processing bias for sexual material: The emotional Stroop

and sexual offenders. Sexual Abuse: *A Journal of Research and Treatment, 16*, 163-171.

doi:10.1177/107906320401600206

Spiegelhalder, K., Kyle, S. D., Feige, B., Prem, M., Nissen, C., Espie, C. A., & Riemann, D.

(2010). The impact of sleep-related attentional bias on polysomnographically measured

sleep in primary insomnia. *Sleep: Journal of Sleep and Sleep Disorders Research, 33*,

107-112.

Stinson, J. D. & Becker, J. V. (2008). Assessing sexual deviance: A comparison of physiological,

historical, and self-report measures. *Journal of Psychiatric Practice, 14*, 379-388.

Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the

Static-99 in practice. *Sexual Abuse: A Journal of Research and Treatment, 24*, 289-302.

doi:10.1177/1079063211423943

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*

Psychology, 18, 643-662. doi: 10.1037/h0054651

Tong, D. (2007). The penile plethysmograph, Abel assessment for sexual interest, and MSI-II:

Are they speaking the same language? *American Journal of Family Therapy, 35*, 187-

202. doi: 10.1080/01926180701226762

Truman, J., Langton, L., & Planty, M. (2013). *Criminal Victimization, 2012*. U.S. Department of

Justice, Bureau of Justice Statistics.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and

psychopathology. *Psychological Bulletin, 120*, 3-24. doi: 10.1037/0033-2909.120.1.3

Wilson, G. D. (1978). *The Secrets of Sexual Fantasy*. London: J.M. Dent.

Worling, J. (2011). The assessment and treatment of deviant sexual arousal with adolescents who have offended sexually. *Journal of Sexual Aggression*, 1-28**.** doi:10.1080/13552600.2011.630152

Wright, L. & Adams, H. (1994). Assessment of sexual preference using a choice reaction time task. *Journal of Psychopathology and Behavioral Assessment, 16*, 221-231. doi: 10.1007/BF02229209

Wright, L. W. & Adams, H. E. (1999). The effects of stimuli that vary in erotic content on cognitive processes. The Journal of Sex Research, 36, 145-151. doi: 10.1080/00224499909551979

Zgoba, K. M., & Simon, L. M. J. (2005). Recidivism rates of sexual offenders up to 7 years later: Does treatment matter? *Criminal Justice Review, 30*, 155-173. doi:10.1177/0734016805284146