Viewing Time and Choice Reaction Time: Exploring its Utility with Child Sex Offenders in New Zealand

A thesis presented in partial fulfilment of the requirements for the degree of Master of Arts in Psychology at Massey University, Albany, New Zealand

Sheila Ayala Silva

2011

ABSTRACT

This study explores the utility of Viewing Time and Choice Reaction Time in the assessment of child sex offenders in New Zealand. The assessment of sexual interest remains a challenge for clinicians working with child sex offenders. Child sex offenders are less likely to disclose sexual interest towards children for fear of potential repercussions and these issues become evident when offenders attend treatment for their offending. The efficacy of treatment depends upon having reliable information on the individual's sexual interest. The Plethysmograph assessment has been the most widely used assessment of sexual arousal, but research into the use of alternative assessments that are less intrusive is needed. Participants in this study were 52 child sex offenders who were attending treatment at Te Piriti Special Treatment Unit. Participants were assessed on two occasions with the VT and CRT assessments while they were attending the preparatory stages of the programme. Results indicates that response times are not reliable over time and that these assessments cannot identify child sex offenders according to their level of sexual deviance as determined by the STABLE-2007, or the gender and age of the victims they offended against. The findings, possible explanations, and limitations for this study are discussed and recommendations are given for future research.

ACKNOWLEDGEMENTS

First of all, I would like to thank Jim van Rensburg, Principal Psychologist at Te Piriti for his on-going support and belief in me. It has been a long journey and I will be eternally grateful to you.

A big thanks to David Jones, Senior Psychologist at Te Piriti for giving me the idea to write this thesis, for getting me started and for his help at the very end. Thank you David.

To Mate Webb, Cultural Consultant at Te Piriti, thank you for sharing your knowledge and support when discussing cultural aspects of this research.

To Aimee Press and Kirsty Blackwood, Senior Psychologists at Te Piriti, a big thanks to both of you for your emotional support, constant encouragement and for being there when I needed you.

To Sheryle-ann Sadiman, thank you very much for all your help and kindness throughout these years.

To my supervisor Mei Williams, thank you very much for your time, your thorough feedback and your support through this time.

I would also like to thank my husband for his unconditional support and constant encouragement to help me through these years. I would not have been able to do this without you. JEDFE.

To my son, you are my inspiration and motivation. Te quiero mucho hijito.

To my parents, gracias por inculcarme buenos valores, por ensenarme a no rendirme, a seguir adelante y a luchar. Los quiero mucho a los dos.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Child sexual abuse is one of the most sensitive problems in society since it involves the most vulnerable members; children. The negative effects that this type of abuse has on victims in the short and long term have been well documented in the literature. In the short term, children who have been sexually abused may experience a range of psychological and behavioural problems, such as early sexualisation, anxiety, depression, aggressive behaviour, difficulties with their performance at school, truancy as well as offending behaviour (Mullen, King, & Tonge, 2000). In the long term, the experience of child sexual abuse has been linked to increased anxiety and depression symptoms (Fergus & Keel, 2005), increased risk of eating disorders, substance abuse and problems with their sexual relationships (Fanslow, Robinson, Crengle, & Perese, 2007).

The prevalence of child sexual abuse in New Zealand has been estimated by several studies. The Otago Women's Health Survey conducted a research between 1989 and 1990 (Mullen, Anderson, Roman-Clarkson, and Martin, 1991). The study found that 32% of 497 women interviewed reported they had been sexually abused before the age of 16. Furthermore, the New Zealand Youth 2000 study gathered information from 10.000 male and female high school students (McGregor, 2008). The study found that 26 percent of female students and 14 per cent of male students had reported unwanted sexual experiences (Fleming et al., 2007). A later study by Fanslow, Robinson, Crengle, and Perese (2007) gathered data from 2,855 women aged between 18 and 64 years old, from both the Auckland and North Waikato. The study found that 24% of women living in Auckland and 28% of women living in North Waikato had experiences of sexual abuse or unwanted sexual contact before the age of 15.

As a result of the impact that child sexual abuse has on victims, families and society in general, governments face increased pressure to protect society from people who pose a risk to the safety of the community. In line with the above, some amendments were made to the Sentencing Act 2002 and Parole Act 2002 in order to create further deterrents for people committing sexual offences against children. The Sentencing Act 2002 implemented a sentence of Preventive Detention, which is an indefinite sentence imposed on offenders who have an on-going risk of committing further offences. The

Parole Act 2002 amendments allowed for a better monitoring of offenders, by means of residential restrictions, home detention and electronic monitoring. In addition, since 7 July 2004 a further Extended Supervision Order can be imposed on high risk prisoners.

Information from the Principal Advisor Policy and Operations from the Department of Corrections, New Zealand, showed that since 1990 the re-offending rate has fallen (A. Skelton, personal communication, February 22, 2011). The re-offending rates fell from 9% to 4% after 2003, which is one year after the law change and new restrictions were imposed on offenders. The number of child sexual offenders sentenced to two or more years however appears to be relatively stable across time, with an average of 200 offenders sentenced every year.

Along with the implementation of the above legislation, emphasis has been placed on identifying those offenders who are more likely to recidivate. Deviant sexual interests have been shown to be predictive of sexual recidivism (Hanson & Bussiere, 1998). However, there are issues around assessment, since child sex offenders are often reluctant to disclose their sexual interests (Hunter, Becker, & Kaplan, 1995; Laws, Hanson, Osborn, & Geenbaum, 2000; Seto, 2008; Seto, 2009). Several measures have been developed in the assessment of sexual deviance. The Plethysmograph assessment is the most used among sex offenders. Despite the important information that this assessment provides it faces several limitations, such as being intrusive, the lack of standardised procedures and equipment, increased costs in terms of specialised training and most importantly it is prone to faking. This study will explore the utility of the newer assessment methods Viewing Time and Choice Reaction Time on a population of New Zealand child sex offenders.

CHAPTER ONE

LITTERATURE REVIEW

Sexual Interest and Sexual Deviance in Child Sex Offenders

Sexual interest, sexual attraction and sexual preference are concepts that appear to overlap in the literature, but are not identical. Price and Hanson (2007) define sexual interests as the thoughts, ideas, beliefs, or preferences that individuals find sexually attractive, stimulating and that can lead them to sexual arousal. Sexual attraction appears to be related to the individual's subjective measure of what they consider attractive in a person and this is closely related to the individual's sexual interest (Levay & Valente, 2006; Rathus, Nevid, & Fichner-Rathus, 2008). While sexual preferences appear to be related to sexual orientation, sexual interest appears to precede the individual's sexual preference, the individual's self-identification as heterosexual or homosexual and also any engagement in sexual behaviour (Seto, 2008).

With regards to sexual deviance, there is no exact definition of sexual deviance in the literature. It appears that sexual deviance is defined from the perspective of behaviours that are considered less "normal" than others and that a value judgment is applied when deciding whether a behaviour is to be considered deviant or not (Laws & O'Donohue, 2008). However, sexual interest in children would be considered deviant in most westernised cultures. It is often assumed that men who offend against children have a sexual interest in children (American Psychiatric Association, 2000), and therefore the focus of treatment providers is on assessing offenders' sexual interests, specifically their deviant sexual interests. Even though sexual interest occur at a cognitive level it is difficult to assess and measure (Price & Hanson, 2007) and therefore child sex offenders are often assessed from the perspective of the individual's sexual arousal in response to sexual stimuli, such as in phallometric assessments. This reflects what Rupp and Wallen (2008) indicated as an unclear independence of the cognitive, emotional and

physiological aspects of sexual arousal and that each of these aspects influence each other in producing a response to the sexual stimuli.

Sexual preference towards children has also been defined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Revised (DSM-IV-TR; American Psychiatric Association, 2000). Paedophilia is included under Paraphilias and "involves sexual activity with a prepubescent child (generally age 13 years or younger). The individual with Paedophilia must be age 16 years or older and at least 5 years older than the child" (American Psychiatric Association, 2000. p. 571). The urge or arousal towards children should be present for at least 6 months. In addition the sexual urges cause distress and difficulties in the individual's interpersonal functioning (American Psychiatric Association, 2000).

The DSM-IV-TR's definition has raised some issues. Grubin (2008) considered that a paedophile is sexually deviant, but that this didn't warrant the person having a mental disorder. In addition, Marshall (1997) indicated that the lack of distress or interpersonal difficulties did not mean that a person could not be considered as having a diagnosis of paedophilia. Furthermore, there is current debate on whether the fifth edition of the DSM should establish a differentiation between people who are sexually attracted to pre-pubescent children and those who are attracted to pubescent children (Blanchard, 2009). Paedophilia will still apply to the former group, while hebephilia will refer to the latter. This distinction may be relevant when trying to differentiate paedophiles from non-paedophiles, since paedophiles appear to prefer children who show no signs of secondary sexual development, which usually start during puberty (Seto, 2008). Therefore, it is argued that this difference will provide a more clear understanding of the individual's sexual preference and potentially aid in their treatment (Blanchard et al., 2009).

The main issue surrounding the inclusion of paedophilia in the DSM-V is related to whether paedophilia should be seen as a mental disorder. Marshall and Eccles (1991) had already pointed out that the diagnostic criteria of paedophilia in the DSM have had implications for the treatment and risk assessment of child sexual offenders. They argued that the diagnostic criteria excluded many chronic sex offenders from it. Similar

issues were discussed in a more recent article by Blanchard (2009). This debate is yet to be resolved.

The term paedophilia is often used when referring to people who have sexually offended against children. This statement implies that child sex offenders in general are paedophiles. However, this may not necessarily be true for all child sex offenders (Seto, 2008).

Seto (2008) suggests that paedophiles and non-paedophiles may have similar curiosities in terms of child nudity while they are children themselves, but that paedophiles remain fixed at this stage and may not develop a sexual interest in adults as they mature. It was suggested that the main feature of paedophiles is the persistence of sexual interest in children. Distinguishing between true paedophiles and those who offend against children but who are not paedophiles are important for researchers and treatment providers.

In line with the above, some studies have suggested a possible differentiation between child sex offenders and paedophiles. Groth and Birmbaum (1978) suggested that there were two types of child sex offenders; fixated offenders and regressed offenders. Fixated offenders were ones who showed greater sexual interest in children and whose victims were usually unknown and non-related. On the contrary, regressed offenders sexually preferred adults to children but due to certain circumstances used children as their sexual partners, and their victims were usually from within their own family. More recently, Holmes and Holmes (2002) provided a typology of sexual offenders. They divided child sexual offenders into situational offenders, those who sexually offended against children under certain circumstances but did not have an actual sexual preference for children, and preferential offenders whose primary sexual preference were children. However, the latest typology was criticised due to the limited information regarding the number of cases involved for Holmes and Holmes (2002) to draw their conclusions on the typology of sexual offenders (Cox, 2006).

Identifying whether an individual has a deviant sexual interest is important due to its relationship with sexual recidivism. An explanation of their association will follow next.

Sexual Deviance and Recidivism

Research has shown that deviant sexual arousal remains an important predictor of sexual recidivism among child sex offenders (Hanson & Bussiere, 1998; Hanson & Morton-Bourgon, 2004), and therefore understanding how a person becomes sexually deviant and how to assess deviant sexual preferences becomes essential in the prevention and treatment of sexual offending.

Freund and Watson (1992) noticed that sexual offenders who had more than two victims, who were not related to them, showed greater sexual arousal to phallometric assessment stimuli involving children rather than adults. In addition, Abel et al., (1987) found that offenders who offended against unrelated male children had more victims than those who offended against female children. Freund and Watson (1992) suggested that because those who offended against unrelated male children were more likely to have more victims, this group of offenders may represent true paedophiles.

Hanson and Bussiere's (1998) meta-analysis of studies of sexual offender recidivism found that factors related to sexual deviance were predictors of sexual recidivism. Those who were identified as being more deviant as measured by the phallometric assessment or on self-report measures were offenders who had more than one child victim, male children victims, children that were unknown to the offender, and showing sexual arousal to child stimuli. Hanson and Morton-Bourgon's (2004) updated review of their 1998 study confirmed that offenders who were found to have deviant sexual interests were more likely to continue their sexual offending. Thus it would appear that having reliable and valid assessment measures of sexual deviance would be an important aspect in treatment and prevention of child sexual offending.

Explanations for Sexual Deviance and Child Sexual Offending

Single Factor Theories of Child Sexual Offending

Several studies have focused their attention on explaining how sexually deviant interests emerge. Some of these explanations are described below.

McGuire, Carlisle, and Young's (1965) study found that deviant sexual arousal was the product of a gradual process of conditioning that involved the pairing of deviant fantasies and masturbation to those fantasies. In addition, the literature suggests that these deviant preferences precede deviant acts and that they are acquired before the age of 18 years (Abel et al., 1987; Marshall, Barbaree, & Eccles, 1991).

Attachment has also been considered to play a role in the development of sexual deviance (Ward, Hudson, Marshall, & Siegert, 1995). Attachment refers to the bond between the child and parent. A positive bond will allow the child to feel secure and confident to establish adequate relationships with peers. In contrast, a person who grew in an environment that did not allow him to feel safe and secure, and only provided him with negative experiences such as exposure to family violence, dysfunctional family environments, early exposure to pornography, sexual and/or emotional abuse may lack the skills to be able to relate to peers (Wieckowski, Hartsoe, Mayer, & Shortz, 1998). In this context, such persons may look for ways to compensate for their interpersonal and emotional deficits by relating to younger people, since they are less threatening and they will be less likely to judge them (Lussier, Beauregard, Proulx, & Nicole, 2005).

In line with the above, some other theories have provided some explanation into the development of deviant sexual behaviour. For instance, some studies have suggested that the birth order of a child sex offender may be linked to deviant sexual preferences (Lalumière, Harris, Quinsey, & Rice, 1998). This is explained from the perspective of an early disruption on the sexual differentiation of the male foetal brain due to maternal immunoreactivity during pregnancy. This means that the more male children a woman gives birth to the bigger the probability becomes that the youngest child will be affected by the mother's immune response (Lalumière et al.,1998). However, a later study showed that although there is some evidence that supports this theory, further studies are needed with larger samples (Côté, Earls, & Lalumière, 2002).

The role of sexual fantasy in sexual deviance has also been explored. McGuire, Carlisle, and Young (1965) found a relationship between deviant fantasies and arousal to those fantasies. Thereafter Leitenberg and Henning (1995) in their study suggested that

masturbation to deviant fantasies played a role in the reinforcement and maintenance of those deviant fantasies. Several procedures were developed to reduce the reinforcement of those deviant fantasies, such as thematic shift (Marquis, 1970), masturbatory satiation (Laws & Marshall, 1991), amongst others. Laws and Marshall (1991) stated that the thematic shift procedure was not considered effective due to confounding of different techniques.

Recent research has focused however on examining whether child sex offenders used deviant fantasies before and after their first offense (Dandescu & Wolfe, 2003). The study found that child sex offenders reported having more deviant fantasies after they had first offended. This suggests that deviant fantasies are not fully established before the first offense and that deviant fantasies played a bigger role in the maintenance of deviant behaviour rather than in its aetiology. The study also notes that there was a correlation between the number of months the child sex offenders spent in treatment and an increase in reporting their deviant sexual fantasies. Dandescu and Wolfe (2003) suggest that the above correlation may be due to child sex offenders increased awareness of their deviant fantasies, being more open to talk about them or a potential influence from the treatment model on child sex offenders. A similar correlation was also found in a previous study (Swaffer, Hollin, Beech, Beckett, & Fisher, 2000).

The above theories attempt to provide some explanation into the causes and development of sexual deviance. However, it has been suggested that due to its complexity, child sexual abuse cannot be fully explained by single factor theories. It is therefore that more contemporary and multifactorial theories have been developed for this purpose (Ward, 2002).

Multifactorial Theories of Child Sexual Offending

Four influential theories have been developed to provide an explanation of the multiple factors involved in child sexual offending. These are: Four Preconditions Model of Sexual Abuse (Finkelhor, 1984) , Quadripartite Model of Sexual Aggression (Hall & Hirschman, 1992), Integrated Theory of the Etiology of Sexual Offending (Marshall &

Barbaree, 1990) and The Pathways Model of Child Sexual Abuse (Ward & Siegert, 2002).

Four Preconditions Model of Sexual Abuse (Finkelhor, 1984)

The four preconditions model was developed by Finkelhor (1984) as a way to provide an explanation of child sexual abuse. The model aims to integrate both psychological and sociological factors and also to remain general enough to capture the variety of situations in which sexual abuse could occur. The model suggests that there are four underlying factors that would explain the behaviour of a child sex offender: Emotional congruence, which refers to the offender's belief that sexual contact with children will give him emotional gratification; Sexual arousal, which refers to offenders feeling sexually aroused by a child; Blockage, which refers to the offenders' inability to meet their emotional and sexual gratification with adults; and Disinhibition, which refers to offenders not been dissuaded by norms to offend sexually against children. The above factors are seen as contributing to one of four preconditions that must be met for child sexual abuse to occur.

The four preconditions of this model are:

1. Offender's motivation to sexually abuse
2. Means of overcoming internal inhibitors
3. Means of overcoming external inhibitors
4. Means of overcoming the resistance of the child

Finkelhor's theory was the first multifactorial theory to be developed. It provided a basis for the study of men who sexually offend against children and it also offered a framework for the use of this model in clinical interventions (Finkelhor, 1984). Despite the inclusion of psychological and sociological factors in the model, developmental and cognitive factors were not considered. In addition, it was later criticised for its apparent contradiction in terms of whether sexual abuse is caused by one single factor or an interaction of factors (Ward & Sorbello, 2003)

Quadripartite Model of Sexual Aggression (Hall & Hirschman, 1992)

Hall and Hirschman's (1992) model was initially developed to understand men who offended sexually against adults, before it was extended to understand sexual abuse against children. Hall and Hirschman's four component theory involved physical sexual arousal, cognitive distortions, affective dyscontrol, and personality factors. According to this theory early negative experiences, such as offenders' own experience of sexual abuse, may lead to lasting personality problems and consequently may increase their likelihood of offending sexually against children.

In this theory the interaction of the four components can potentially increase the likelihood of committing a sexual offense against a child. However, one of those components alone could still be sufficient for an offender to offend. This one component thus becomes the offender's primary motive to offend. Furthermore, different combinations of the various components relate to different offender types with specific treatment needs. This is a more specialised approach than what Finkelhor presented in his model (Finkelhor, 1984). Hall and Hirschman's quadripartite model introduced the notion of multifactors and acknowledged individual differences; however it was considered that a more integrated model that could explain complex aspects of child sexual offending was more pertinent (Marshall & Barbaree, 1990)

Integrated Theory of the Etiology of Sexual Offending (Marshall & Barbaree, 1990)

The Integrated Theory proposes that child sexual offending occurs as a consequence of interacting developmental and environmental factors (Marshall & Barbaree, 1990). According to this theory, the transition into adolescence becomes a critical period for the development of enduring sexual attitudes and behaviours. Marshall and Barbaree (1990) pointed out that sex and aggression come from the same neural substrates. Early experiences and environmental factors play an important role in the manner in which the individual learns to express his sexual needs and maintain control over his aggression. It is at the time of puberty where the young male will be faced with the challenge to learn to separate sex and aggression and to learn to refrain from aggression within a sexual context.

Experiences such as physical, sexual and/or emotional abuse makes it more difficult for the young male to prepare and to adjust to the physical changes experienced during puberty and potentially leading him to a strong desire to engage in sex and aggression. Due to these early experiences he may feel resentful, hostile and lacking self-confidence to form peer age relationships and therefore may look for alternative ways to still fulfil his emotional needs, such as contact with children. Because of the emotional indifference they themselves experienced while growing up, they may become indifferent towards others and lack the empathy to prevent them from hurting other people (Marshall & Barbaree, 1990).

In addition, environmental factors such as messages sent by society and the media about sex and aggression, the availability and access of pornography as well as the circumstances surrounding the sexual offences (alcohol use, anger, emotional dysregulation) are all considered contributing factors for the lack of separation of sex and aggression, making the individual likely to commit a sexual offense (Marshall & Barbaree, 1990).

The Integrated Theory indicates that sexual offending against children is much more complex and requires further exploration into early experiences and environmental factors in order to understand its development and occurrence. The fusion of aggression and sex described in the Integrated Theory was considered to be problematic, since aggression is not necessarily present in all sexual offences committed against children (Ward & Sorbello, 2003). The Integrated Theory provided some distinction among child sex offenders; however this distinction became more specific when Ward and Siegert (2002) developed the Pathways model.

The Pathways Model of Child Sexual Abuse (Ward & Siegert, 2002)

Ward and Siegert (2002) proposed that the Pathways model could provide a more comprehensive explanation of adult men who sexually offend against children. This model integrated the major strengths of the three theories explained above and proposed that child sex offenders are heterogeneous. The pathways model recognises that each individual follows a different path to sexual offending. These pathways are influenced

by an interaction of biological, psychological, environmental, emotional, physiological and cultural variables.

In addition, the model suggests that there are four interacting psychological mechanisms (vulnerability factors) that explain sexual offending against children: intimacy and social skill deficits, distorted sexual scripts, emotional dysregulation, and cognitive distortions. The model suggests that the above four mechanisms always interact for an individual to commit an offence, however each individual will have a pathway that will determine the core deficits and these deficits will have an impact on the other psychological mechanisms. There are five pathways in this model.

Intimacy Deficits Pathway

This pathway refers to individuals whose sexual preference is for adults, but under situations in which their partner is unavailable they will look to meet their needs for closeness and intimacy with a child. The core cause underlying this pathway appears to be related to insecure attachment. Offending will tend to occur after the offender has been through a period of feeling lonely and rejected and feeling unsatisfied with his relationships. Offending often starts in adulthood and the offender's self-esteem may be high or low depending on the individual (Ward & Siegert , 2002)

Deviant Sexual Scripts Pathways

This pathway refers to individuals who have developed distorted sexual scripts about sex and intimacy. One possible explanation for this is that these men were sexually abused themselves and became early sexualised. Their sexual preference is still for adults; however their inability to separate sex and closeness may lead them to experience unsatisfying relationships with adults. Child sexual offending may occur if there is a child present, who is perceived by the offender as being less threatening, more trustworthy and a mean to meet his sexual and emotional need. These men will often start offending as adults and will present with low self-esteem due to their experiences of rejection and loneliness in their relationships (Ward & Siegert , 2002)

Emotional Dysregulation Pathway

The model suggests that individuals in this pathway have normal sexual scripts but lack the ability to regulate their emotions and look at sex as a way to cope with emotional distress. This means that these men will have difficulties identifying and managing their negative emotions and in conjunction with sexual arousal, it may lead individuals to become disinhibited or, given the opportunity, may offend against a child. Because these men have normal sexual scripts, their preference and sexual fantasies are about adult relationships, but when they find themselves under emotional stress they are more likely to offend against a child. These offenders may start offending during adolescence or adulthood and their self-esteem will vary depending on their early experiences and backgrounds (Ward & Siegert , 2002)

Antisocial Cognitions Pathway

Men within this pathway will have normal sexual scripts; however they will have developed general antisocial attitudes and beliefs. Sexual offending against children appears to be related to a general tendency to engage in antisocial behaviour rather than a sexual preference for children. Their antisocial attitudes and beliefs linked to opportunity and sexual desire may lead to child sexual offending. These individuals may find that child sexual abuse is pleasurable, since they are meeting their need for self-gratification (Ward & Siegert , 2002).

Multiple Dysfunctional Mechanisms Pathway

The model indicates that individuals within this pathway will have distorted sexual scripts, which will lead to sexually deviant fantasies and behaviours. They will often have a history of sexual abuse as children and/or the early exposure to pornographic material. The model indicates that these men are likely to engage in deviant sexual fantasies prior to their first offense due to their early experiences of sexualisation. Due to the presence of multiple deficits in their psychological mechanisms and their sexual preference for children, these offenders are considered to be paedophiles. Offending is likely to take place if there is a child present and those offenders who have been able to manage their desire to offend are still considered to be at risk of offending. Even though

deviant sexual fantasies may have started early, offenders may have only started abusing children sexually upon reaching adulthood. These men may have high self-esteem because of their belief that children are sexual partners and therefore that their sexual abuse of children is legitimate (Ward & Siegert , 2002).

The Pathways model provides a structure that guides and facilitates clinical intervention. The model is still a provisional framework that needs further research to support the pathways contained in the model. The focus of the model appears to be in understanding why a man starts offending rather than the maintenance of the offending behaviour. (Ward & Siegert (2002). Connolly (2004) explored whether the Pathways model could be explained with 13 child sex offenders attending the Kia Marama Special Treatment Unit, Rolleston Prison. The study found that the pathways model could fit with 10 of the men that took part in the research. However, the small sample size limited the generalisation of the results.

Overall, each of the four theories mentioned above have contributed to the further understanding of child sexual offending. Finkelhor's theory (1984) set the basis for a multifactorial explanation of child sexual offending, while Hall and Hirschman's quadripartite model (1992) allowed for a more distinct typology among child sex offenders that could facilitate clinical interventions. The Integrated Theory (Marshall & Barbaree, 1990) acknowledged the complexity of child sexual offending and put emphasis on the individual's early experiences and environmental factors in order to understand how child sex offending developed and occurred. Finally, the Pathways model (Ward & Siegert, 2002) seeks to address the limitations of the previous theories. It provides a framework that facilitates the understanding of the aetiology of child sex offending, leading to more specific goals for treatment. However, due to the complex aspects that surround child sexual offending, the Pathways model is still in need of further research and development.

Overview of Current Assessments of Sexual Deviance

In the assessment of deviant sexual interest, phallometry has been considered the best measure available for the assessment of sexual interest (Seto, 2001) Other alternative

assessments have been developed for the measurement of sexual interest, but they have not been as well researched as the phallometry type assessments (Kalmus & Beech, 2005). The measures of sexual interest can be divided into three categories: physiological, non-physiological and attentional assessments. These assessments will be briefly reviewed.

One of the main issues when assessing child sex offenders is their lack of compliance and tendency to fake their results. Offenders will often avoid disclosing any information that could indicate deviant sexual preferences for fear of the consequences of those disclosures (Seto, 2008; Wormith, 1986). It is for this reason that phallometric assessments are widely used, since one of its main advantages is that is not relying on self-reported sexual preference (Earls, 1983). However, the use of this assessment has been controversial due to the intrusive nature of the procedure, the ethical concerns regarding the pictures that are shown, the specialised equipment that is required and the associated costs and length of time needed to train examiners to administer the measure (Craig, Browne, Beech, & Stringer, 2006; Hunter, Becker, & Kaplan, 1995; Kalichman, Henderson, Shealy, & Dwyer, 1992).

Physiological Assessment

Phallometric Assessment or Penile Plethysmograph (PPG)

The Penile Plethysmograph (PPG) was developed in the 1950s by Kurt Freund (1967) for the purpose of detecting men who were claiming to be homosexuals in order to avoid military service. The PPG measures changes in penile responses when individuals are exposed to audio and/or visual material that involve people of different age and gender groups, in contexts that may involve consent or coercion. The images may include people that are clothed or unclothed (Flak, Beech, & Fisher, 2007).

The use of phallometric assessments to measure penile changes to sexual stimuli is related to the third stage of Singer's theory (1984) of sexual arousal. His study stated that "genital changes are the most reliable and convenient to measure" as the

autonomic and somatic components of sexual arousal (Singer, 1984, p.234). Penile responses can be measured either by changes in volume or circumference, and the increases are interpreted as sexual arousal to the stimuli presented. The original volumetric devices were developed by Freund (Freund, 1963). Even though it appears that they may be superior to the circumferential devices (Marshall, 2006), circumferential gauges are the most widely used devices in phallometric assessment since they are easier to use (Kalmus & Beech, 2005; Seto, 2008).

Freund (1965) found that responses to phallometric assessments could differentiate between sex offenders and non-offenders. Some other studies have found similar results (Barbaree & Marshall, 1989; Freund, 1965; Freund, 1967; Quinsey & Chaplin, 1988). With regard to the PPG's discriminative validity, it appears that phallometry can discriminate between child sex offenders and non sex offenders and that the offender's responses to the assessment are associated with the gender of the victim they offended against (Seto, 2008). In relation to its predictive validity, Hanson and Bussière (1998) found that sexual arousal to children on phallometric assessment was considered to be the strongest predictor of sexual recidivism. A later study by Hanson and Morton-Bourgon (2004) supported this.

Phallometric assessments are also known to have several limitations, such as a lack of standardization on the stimuli material, the procedure and the interpretation and analysis of results. Kalmus and Beech (2005) found in their review that the device used to measure penile responses seemed to be the only common thing among the studies. It appears that even though the Association for the Treatment of Sexual Abusers (ATSA) had developed guidelines on the phallometric assessment procedures, several laboratories doing phallometric assessment still showed differences in their methodologies (Merdian & Jones, 2011). This becomes a problem when attempting to develop normative data that will facilitate the interpretation of the results (Seto, 2008), as well as when trying to replicate the findings and compare the studies (Murphy & Barbaree, 1994). Phallometric assessments have also been criticised for its lack of reliability. Few studies have been done on reliability and when reliability was reported it was often below what was expected for a psychometric test (Flak et al., 2007).

A second limitation involves the potential for faking. Offenders may avoid looking at the stimuli presented and/or use distraction techniques to prevent any arousal and this is difficult to control or monitor. In addition, men can manipulate their erectile responses by manipulating their perineal muscle contractions (Laws & Gress, 2004). Phallometric assessments are transparent and offenders are aware that their sexual responses are being recorded and therefore will attempt to suppress arousal to any deviant stimuli and increase it towards a more appropriate stimulus (Merdian & Jones, 2011). Due to the on-going issues with faking in the detection of sexual arousal, galvanic skin response (GSR) was used in two studies in conjunction with the PPG assessment to increase the accuracy of the results (Card & Farrall, 1990; Farrall, & Card, 1988). The studies indicated that arousal was more difficult to fake and that GSR was capable of detecting participants' attempts to suppress arousal and voluntary arousal. However, no further studies have been done on the use of GSR and faking in sexual arousal. Some strategies have been considered in order to reduce faking, such as asking individuals to push buttons when viewing or listening to stimuli of violent or sexual content (Quinsey & Chaplin, 1988). Even though, there are still no procedures to control the frequency of faking, there is some confidence that control methods have improved the effectiveness of PPG (Marshall, 2006; Merdian & Jones, 2011).

Low responders and non-responders have also been an issue in phallometric assessments. Kalmus and Beech (2005) noted that there was a tendency to exclude low responders; however this could potentially misrepresent group differences if only high responders were entered into the analysis. Looman, Abracen, Maillet, and DiFazio (1998) found that there was a relation between low responders and Social Desirability Scales, which may indicate that offenders are voluntarily suppressing their arousal. The literature has also shown that phallometric assessments may be more accurate when offenders have admitted to deviant sexual interests rather than denying it (Freund & Watson, 1991). Non admitters usually produce normal arousal patterns (Marshall, 2006).

Another issue with PPG appears to be that when comparing offenders to non-offenders it is assumed that non offenders will not have any deviant interest; however it has been shown that normal men will have a degree of deviant sexual interest (Marshall, 2006;

Merdian & Jones, 2011). Ethical issues regarding the intrusiveness of the measure and the impact that stimuli may have on the individual, such as sexually abused offenders, are also issues to consider when administering phallometric assessments (Merdian & Jones, 2011).

Overall, PPG have several limitations, but it is still the most widely used for the assessment of sexual preference. Research into alternative measures of sexual interest that could address or minimise those limitations is therefore warranted.

Thermistor or Thermography

Thermistor devices were introduced by Abramson, Perry, Seeley, Seeley, and Rothblatt (1981) with the purpose of analysing temperature changes during sexual arousal. The theory behind the thermistor suggests that it can detect changes in vascular and blood flow related to changes in skin temperature (Abramson, Perry, Seeley, Seeley, & Rothblatt, 1981) Two studies found a correlation between thermistor, self-report, and plethysmography data on non-sexual offenders (Abramson et al., 1981; Webster & Hammer, 1983). However, Beck, Barlow, and Sakheim (1983) found that the thermistor was not sensitive enough to detect different levels of arousal. In addition, one of the disadvantages of this technique is that it requires highly specialised equipment (Beck, Barlow, & Sakheim, 1983)

Photoelectric Surface Blood Volume Measurement

This technique is based on the premise that when the blood flow in the tissues of the penis increases during arousal, then the blood volume in the surface also increases (Kalmus & Beech, 2005). The photoelectric surface blood volume was used with twenty four sexual offenders from a maximum security hospital who were under assessment (Hinton, O'Neill, & Webster, 1980). The participants were shown films that included interactions of consent and the use of force with different gender and age groups (Hinton et al., 1980). Self-report data, Plethysmograph results, and the measurement of the participants' surface blood volume were compared. Even though this assessment appeared to have been able to differentiate between non sex offenders and sex

offenders, a high number of responses appeared to have been the result of participants' defensiveness and faking. There doesn't seem to be further studies on the use of this technique.

Pupillometry

Pupillometry was first developed by Hess and Polt (1960) and it is based on the premise that the individual's pupil will dilate when exposed to stimuli that they find sexually interesting and that the pupil will constrict when is not. However, Hess and Polt's study has been criticised for the use of small sample size (Janisse, 1977; Zuckerman, 1971), and issues with the statistics they chose (Janisse, 1977). No support has been found for Hess and Polt's theory that dilation was related to positive sexual interest and that contraction reflected the opposite (Janisse, 1977; Scott, Wells, Wood, & Morgan, 1967; Zuckerman, 1971). In addition issues with the use of illumination while the subjects were assessed have also been raised as having an impact in pupil dilation (Aboyoun & Dabbs, 1998; Goldwater, 1972; Janisse, 1977; Zuckerman, 1971). Although Hamel (1974) showed that there was a potential relation between pupillary dilation and sexual interest, there have not been further studies on pupillometry and its ability to reliably differentiate sexual preference (Garrett, Harrison, & Kelly, 1989).

Electroencephalography

The aim of this type of assessment is to measure a neurological effect of anticipation when the individual is aware that he is going to be exposed to his preferred visual sexual stimuli. This technique could potentially address some of the issues related to faking, since the individual will be less likely to control his responses (Kalmus & Beech, 2005). Howard, Longmore, Mason, and Martin (1994) showed a correlation between the results of this technique and the stated preference of heterosexual and homosexual individuals. However, this type of assessment did not provide a significant difference when paedophiliac offenders were assessed on their sexual preference. Further studies are required for the use of this technique as a reliable measure of sexual preference.

Nonphysiological Assessment

Self-report

Sexual offenders' self-report can be assessed through clinical interviews and questionnaires. Self-report involves asking individuals directly about their sexual thoughts, urges, fantasies and behaviours (Seto, 2009). Even though self-report measures have several advantages with regard to low costs, ease of administration and non-invasive nature in comparison with physiological measures, it poses challenges to clinicians (Holland, Zolondek, Abel, Jordan, & Becker, 2000). Several studies have shown that offenders may be less willing to disclose their deviant sexual preferences due to potential legal repercussions (Hunter, Becker, & Kaplan, 1995; Laws, Hanson, Osborn, & Geenbaum, 2000; Seto, 2008; Seto, 2009). Nevertheless, self-reports still provide relevant information into the individual's sexual motives and behaviours. Seto (2009) suggests some strategies that could improve the gathering of information when interviewing an offender, such as adopting a non-judgmental tone and asking open ended questions in order to encourage disclosure and reduce denial. These strategies are especially useful when doing clinical interviews.

Clinical Interview

Clinical interviews are usually considered the most important part of the assessment process. The purpose of the clinical interview is to gather information from the individual's personal history in order to understand the individual's psychological functioning (Flak et al., 2007). Despite the useful nature of clinical interviews, Abel (1985) found that offenders are less likely to disclose other deviant sexual behaviours rather than their index offences. Beech (2001) pointed out some guidelines that could assist clinicians when conducting interviews, such as maintaining a natural flow during the interview. However, it is acknowledged that interviewing sex offenders may be more difficult, since they have the tendency to deny or dissimulate any deviant sexual preferences. Unless clinicians are able to gather some collateral information, they will have to rely on the offender's self-report. One way of obtaining collateral information is assessing the offender with self-report questionnaires. Abel (1985) found that offenders

are more likely to disclose deviant sexual thoughts and behaviours on self-report questionnaires than to the clinician during a clinical interview.

The Sexual Interest Card Sort Questionnaire

The Sexual Interest Card Sort questionnaire is a self-report measure which includes several Categories of sexual interest. The individual is asked to rate their sexual interest on each of the stimulus describing the categories. The stimuli presented can include either pictures or descriptions related to the sexual categories (Seto, 2008).

Laws, Hanson, Osborn, and Greenbaum (2000) compared the responses of child sex offenders to the Card Sort questionnaire, PPG slides and PPG audio. The results showed that the Card Sort questionnaire had a higher accuracy in discriminating the gender preference of child sex offenders, whether it was a female victim or a male victim. In addition, the study showed that the three measures in combination provided the highest accuracy of classification of child sex offenders. However, due to participants being an exceptional group of admitters of sexual deviant behaviours, the findings of the study need to be treated with caution. It is suggested that the Card Sort questionnaire may be more useful when administered early in the assessment process, due to offenders being more willing to discuss their sexually deviant behaviours at this stage (Abel, 1985).

Holland et al.,(2000) noticed that some offenders found some of the statements mentioned in the Card Sort questionnaire offensive and were reluctant to complete the questionnaire. They then explored whether a shortened version, without the offending statements, could still maintain its psychometric properties. The shortened version could also classify child sex offenders and offenders with several paraphilias, however this study also highlighted that this measure works best when offenders admit to having deviant sexual preferences. Hunter, Becker, and Kaplan (1995, as cited in Kalmus & Beech, 2005) pointed out that when the Card Sort questionnaire was used with a group of young offenders, who were less willing to discuss their deviant sexual behaviors, there was little correlation between their results to the questionnaire and their phallometric data.

The main issue with the Card Sort questionnaire appears to be that it works best when the offenders are willing to disclose their deviant sexual preferences. This raises the issue that the questionnaire is susceptible to denial and faking and therefore is often recommended to be used in conjunction with other assessment measures (Flak et al., 2007).

The Multiphasic Sex Inventory (MSI)

The MSI is a psychometric test designed to assess the psychosexual characteristics in sexual offenders. The MSI consists of 300 questions with a true/false response format, which are divided into 20 scales, including six validity scales (Holland et al., 2000). In addition it provides an index of the offender's sexual history, knowledge and his level of motivation to engage in treatment. The MSI contains three deviance scales: Child Molest, Rape and Exhibitionism. The Child Molest scale includes only offenders who have groomed their victims for long period of times, and not offenders who would offend against strangers (Kalmus & Beech, 2005).

Kalichman, Henderson, Shealy, and Dwyer (1992) evaluated the psychometric properties of the MSI and reported moderate to high reliability; however it appeared that there was some response bias and that the face validity of the test allowed for faking and denial to occur. The MSI has been shown to differentiate between sexual offenders and non-offenders; however there is no available data on the MSI ability to differentiate among sex offenders (Kalmus & Beech, 2005). The MSI is a test that has been used internationally, however due to the norms being American the interpretation of the MSI results need to be treated with caution. This is because many of the questions contained in the MSI are negatively worded and many of the terms used in the MSI are reflective of an American population (Flak et al., 2007).

Dowling, Smith, Proeve, and Lee (2000) compared Australian sex offenders with American sex offenders with the purpose of developing a preliminary set of norms on the MSI for the Australian population. The study found that the MSI profiles of both American and Australian offenders were similar and that both offender populations suppressed their sexual interest in children and had a tendency to deny their sexual interests. Craig, Browne, Beech, and Stringer (2006) used the MSI to assess whether it

can predict sexual reconviction in sexual offenders. The results show that the MSI could potentially aid in the prediction of sexual reconviction. In addition, this study supports Simkins, Ward, Bowman, and Rinck (1989) findings that the MSI could also predict treatment outcome. As a measure of psychosexual characteristics in sexual offenders it appears that the MSI is most effective in detecting when an offender is faking or denying their deviant sexual interests, but it doesn't seem to be as effective when having to differentiate among offenders (Kalmus & Beech, 2005).

Attentional Assessment

The Emotional Stroop Test (EST)

The EST involves measuring the automatic associations that individuals have toward specific stimuli. In the test participants are shown words in different colours and are asked to quickly name the colour of the word while ignoring the actual meaning of the word itself. The EST is based on the premise that individuals will take longer to respond to the test when the stimuli presented is highly salient for them. The time taken to the individual to name the colour from the time the stimuli was presented is considered to be the measure of interference in automatic processing (Stroop, 1935).

Smith and Waterman (2004) modified the Stroop task to include words that will reflect sexual offences. Their sample included a group of sex offenders, violent offenders, non-violent offenders and undergraduate students. This study showed that both sexual and violent offenders were significantly slower to colour name words that related to sexual offending and aggression. Price and Hanson (2007) replicated the above study with a bigger sample and less questionnaires. The results indicated that even though there were differences similar to the ones found in Smith and Waterman's study, the effects were small and the differences between the groups were not statistically significant. One of the main issues with the test appears to be that the word lists used in the test didn't seem to relate to the thoughts and motives of the offenders.

Van Leeuwen (2009) examined cognition biases on a group of paedophiles. The test consisted of participants categorizing words as sexual or non-sexual on superimposed

images of children and adults. Results indicated that the group of paedophiles were faster in categorizing sexual words when they were superimposed on child images. In addition, the results could discriminate sex offenders from non sex offenders. Despite the encouraging results, larger samples are required. Furthermore, this study calculated scores using bias scores which appear to have produced poor test-retest reliabilities in the past. This means that the results need to be interpreted with caution (Snowden, Craig, & Gray, 2011).

Implicit Association Test

The IAT was developed by Greenwald, McGhee, and Schwartz (1998). The IAT involves a series of categorization tasks in which participants are asked to respond rapidly with the help of two response keys. For example, on the first task participants are asked to use the right hand key for items that represent one concept and one category attribute such as 'flower' and 'pleasant' respectively and the left hand key to categorise another concept and category such as 'insect' and 'unpleasant'. The second task involves a switch of those categories and concepts such as right hand key for 'flower' and 'unpleasant' and left hand key for 'insect' and 'pleasant'. It is assumed that the subject will take longer time to respond to concepts and attribute categories that are generally considered incompatible such as 'insect' and 'pleasant'. The difference in the time taken by the subject to respond to each group is recorded and the difference in mean responses will indicate the IAT effect, which relates to the strength of the association between the target category and the attribute that the individual holds. (Hofmann, Gawronski, Gschwendner, Le, & Schmidt, 2005). In 1998, the IAT showed some ability to identify underling attitudes; however it appeared that when the attitudes were related to sensitive topics, such as race or ethnicity participants were less willing to openly disclose their views (Greenwald, McGhee, & Schwartz, 1998).

Throughout the years, the IAT has been adapted to assess different constructs, such as self-esteem and self-concept with attributes such as pleasant and unpleasant (Greenwald & Farnham, 2000). However, there has been some debate regarding the underlying cognitive processes of the IAT results. Some of the suggestions have been that the IAT measures environmental associations (Karpinski & Hilton, 2001), that it measures

salience asymmetries that are familiar to the individuals (Rothermumd & Wentura, 2004) and current research indicated that IAT measures extrapersonal associations (Han, Czellar, Olson, & Fazio, 2010; Han, Olson, & Fazio, 2006; Olson & Fazio, 2004).

Few studies have been done using the IAT with sex offender populations. Mihailides, Devilly, and Ward (2004) found that IAT was able to differentiate between sex offenders and non sex offenders on their cognitive distortions, with child sex offenders showing stronger cognitive distortions related to children. A year later, Gray, Brown, MacCulloch, Smith, and Snowden (2005) compared child sex offenders with non sex offenders on their association of children with sex. The study found that the above association was stronger for child sex offenders. In addition, Nunes, Firestone, and Baldwin (2007) showed that child sex offenders viewed children as more sexually attractive than adults compared to non sex offenders. Although the studies showed encouraging results, one common issue among them appears to be that the sample size was small, and therefore the results cannot be generalised. More recent research found that when comparing the danger of relapse for non sex offenders with that of paedophiles, the latter group showed greater association between erotic with children versus erotic with women **(**Steffens, Yundina, & Panning, 2008). A limitation of this study appeared to be that the above classification was not based on clinical or diagnostic information of offenders, but on a subjective classification made by the therapists. Due to this prior knowledge, the study was able to show the incremental validity of the IAT measure and therefore it is suggested that the study's results are viewed with caution.

IAT has also been used in conjunction with self-report and viewing time assessments in order to create a test that will provide reliable and sound information regarding deviant sexual interest in child sex offenders (Banse, Schmidt, & Clarbour, 2010). However, the study indicated that viewing time assessments were more reliable and had better convergent and discriminant validity than the IAT measures, which only showed satisfactory reliability and/or moderate to low reliability.

Overall, IAT appears to be a measure that could potentially provide information in the assessment of deviant sexual interests, however further research is needed before the IAT can be used confidently and there is some concurrence into how the IAT results

should be interpreted. In addition, viewing time assessments appear to provide better reliability and validity (Schmidt et al., 2010), and therefore it is the measure that will be used for the present research.

Viewing Time

Viewing Time (VT) assessments are based on the premise that individuals will spend longer time looking at pictures that they find sexually attractive and a summary profile of the individuals' viewing time will allow for the differentiation of each individual's sexual interest (Laws & Gress, 2004). It is therefore possible to measure sexual interest to various stimuli presented to a subject.

VT was first used by Rosenzweig (1942) for the measurement of sexual interest among 10 schizophrenic individuals who had been identified as having either high or low interest in sexual activity. An instrument, which was called a photoscope, was developed for the specific purpose of the study. This device was a closed box with a glass window in front. Participants could look at the images through this glass window and could change the pictures they were looking at by pressing on a small knob that was located on the side of the glass window. A stopwatch was used to determine the time participants spent looking at a picture. Participants were shown 24 black and white pictures with a dark grey background, among them five of the pictures involved romantic scenes and five of them involved sexual situations. The results indicated that there was a positive correlation between interest in sexual activity and the time spent looking at the images shown (Rosenzweig, 1942).

VT has been used since Rosenzweig's (1942) study to measure sexual interest in different groups of people. Zamansky (1956) found that homosexual men spent more time looking at pictures of males rather than neutral pictures or pictures of women, which indicated that VT could discriminate people's sexual preferences. Brown, Amoroso, Ware, Pruesse, and Pilkey (1973) evaluated the viewing time of forty male students to pornographic material. The study showed that the viewing times increased when the subjects were alone looking at the stimuli and that the viewing times of the material decreased in the presence of observers. In addition, Love, Sloan, and Schmidt

(1976) found that participants who experienced high degrees of sex guilt showed less interest in sexual stimuli and were less willing to view pornographic material, compared to participants who experienced only moderate degrees of sex guilt feelings.

More recently, Israel and Strassberg (2009) found that VT could discriminate between heterosexual men and women in terms of their sexual interest. Men looked longer at pictures of females and women looked longer at pictures of males. In addition, the study showed that in comparison to women, men's sexual interest is more category-specific. This means that when men are looking at preferred stimuli they will show high level of sexual interest, while they will show low interest when looking at non preferred stimuli. The above studies suggest that VT is a measure that could provide relevant information in the measurement of sexual interest.

Currently Abel Screening Incorporated offers three VT procedures that are commercially available and exclusively marketed: The Abel-Blasingame assessment system for individuals with intellectual disabilities™, previously called Affinity, The Abel Assessment for sexual interest –2™ and the Abel Assessment for Sexual Interest-3™ (AASI-3; Abel Screening Incorporated, 1995).

The Abel-Blasingame Assessment System has been developed to assess the sexual interest of learning-disabled people (Glasgow, 2009). A 2003 study showed that this assessment provided reliable information on the participants' sexual preference and that this assessment could provide supplementary information to self-report measures (Glasgow, Osborne, & Croxen, 2003). The same assessment was used in conjunction with two self-report measures with a sample of adolescent sex offenders (Worling, 2006). The findings showed that the internal consistency of the three assessments in relation to the age and gender groups was acceptable and that the three measures appeared to discriminate between adolescents who had offended against male child victims and those who had not. However, none of the three measures could specifically identify those adolescents who had female victims only. The study attributed the results to offenders having a specific sexual interest toward male children when offending against them, but that this is less well-defined when the offending is against female children (Worling, 2006). The study pointed out that even though VT can potentially

identify sexual interest towards male children in adolescents, individuals' sexual preferences and interests will be better defined in adulthood. Another important aspect of this study is that the participants did not find the assessment upsetting, which emphasises the utility of VT as a less intrusive approach.

The Abel Assessment for Sexual Interest (AASI) has received different names and has served different purposes throughout the years, the Abel Screen, the Abel Interest for Paraphilias and the Abel Assessment for Sexual Interest-2™ (AASI-2; Abel Screening Incorporated, 1995). The AASI-3 is the latest and current version. However, there was no independent research of the AASI-3 available at the time of this study and therefore the review of this test will focus on the previous version, the AASI-2. This assessment is composed of two parts. The first part includes a questionnaire that explores a range of sexual behaviours. The second part consists of the measurement of the individual's objective sexual interest while viewing 160 digital pictures of both male and female children, adolescents and adults who are dressed in bathing suits (Abel, 2000).

Earlier studies assessed the reliability and validity of the AASI (Fischer & Smith, 1999, Harris, Rice, Quinsey, & Chaplin, 1996; Smith & Fischer, 1999). These studies raised concerns regarding the utility of the measure within adults and adolescent populations and its discriminative ability to differentiate between child sex offenders and non sex offenders. However, the main issues with the above studies appeared to be the sample selection and a misconception of what viewing times measured, which is sustained visual attention. The misconception was to consider VT as a diagnostic tool to determine whether a person has sexually offended against a child (Abel, 2000).

In addition, a later study showed that the AASI could differentiate between individuals that were child sex offenders, non-child sex offenders and liar-denier child sex offenders (Abel, Jordan, Hand, Holland, & Phipps, 2001). This means that such a measure could potentially facilitate a more accurate assessment of an individual's sexual interest and consequently risk assessment. However it is noted that even though the study showed the ability to classify individuals into the above three groups, the AASI results should not be seen in isolation from the clinician's overall assessment of the client (Abel et al., 2001).

In the field of child sexual offending, the Plethysmograph assessment is the most widely used assessment in the measurement of sexual interest in children. Therefore comparison of this measure with other assessments of sexual interest has been relevant (Laws & Gress, 2004). With regards to VT, there were some initial concerns that the AASI and Plethysmograph measured two different aspects of sexual interest, meaning that the Plethysmograph assessed a more physiological aspect of sexual arousal, while the AASI could be measuring a cognitive component of sexual arousal (Krueger, Bradford, & Glancy, 1998). However, the literature shows that sexual arousal is a complex process and clear independence of its cognitive, emotional and physiological aspects is not fully established. Instead it is suggested that the above components affect each other in producing a response to sexual stimuli (Rupp & Wallen, 2008)

Some studies compared the AASI with the Plethysmograph (Abel, Huffman, Warberg, & Holland, 1998; Harris, Rice, Quinsey, & Chaplin, 1996; Letourneau, 2002; Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996). The outcome of these studies differed in terms of which assessment discriminated better sexual interest toward a specific stimulus category, such as the female child category being better discriminated by VT. On the other hand, these studies concurred that individuals looked longer at stimuli that included people of their preferred age and gender group.

There are few published studies that assessed only child sex offenders in their sexual interest in children. In 2004, the validity of the Abel Assessment for Sexual Interest ™ (Abel et al., 2004) was explored. Abel et al. (2004) compared the viewing time of adolescent child sex offenders and a control group. The study found that adolescent child sex offenders spent longer time viewing at pictures of children than the control group. In addition, the study also showed that there was a positive correlation between the participants' viewing time, the number of children that the participants had offended and the number of acts of child sex offending the participants' had committed (Abel et al., 2004).

Research into the use of VT as a measure of sexual interest among child sex offenders and normal heterosexual men found that the sexual interest of both groups could be

discriminated by using a deviance index which was calculated from the participants' viewing times (Harris et al., 1996). The study found that child sex offenders looked longer at the slides of children than the non-clinical group. Although the results provide good indication that VT could be used to assess sexual interest, the selection and small sample size used in the study limits the generalization of the results.

Despite the encouraging results of the AASI, the main issue is that the developers have been reluctant to allow outside researchers to investigate the measure's reliability and validity outside of the developers' control and management (Laws & Gress, 2004). The administration procedure, the scoring and the data interpretation of the above assessments is controlled by Abel Screening Incorporated. Once the individual has completed the test, the results are sent to Abel Screening Incorporated for processing and a report on the assessment is sent back to the administrator within 10 to 20 minutes (Williams, 2003). This has encouraged other researchers to develop a separate assessment which measures sexual interest by using viewing time in the same fashion as the Abel Assessment for Sexual Interest-2™ (Laws & Gress, 2004).

One of the main areas that needed to be addressed was the use of standard images for each age category. Despite Abel Screening Incorporated having developed a standard set of images in the VT assessments, there are still some ethical concerns about using images of real people (Gress, 2005). In the 1990s, Konopasky and colleagues (PsyPhy, 1999) developed a set of 64 stimuli that were computer constructed, but still looked like real people. The images contained four age groups, 5, 9, 13 and adult for each gender. There are four images per age group and each of the images appears both clothed and nude (Laws & Gress, 2004).

The constructed images that Konopasky and colleagues had developed (PsyPhy, 1999) were included in the Gress project (Gress, 2001). The purpose of the project was to explore whether computer constructed images used in the VT system could provide similar positive results as when using images of real people. The project involved the recording of the subject's viewing time to each of the images presented. The individual was given a keypad to answer a question located on the right side of the screen. Once the question had been answered the software removed the slide and presented the next

slide. The questions asked were simple, such as “Do you find this person attractive?” (Laws & Gress, 2004). The classifications that were produced by the VT were compared with the ones produced by the Sexual Deviance Card Sort and the offenders’ past sexual behaviour. This study was published in a separate article (Gress, 2005) and will be explained next.

Gress (2005) compared the VT assessment with computer generated images with the Sexual Deviance Card Sort and past sexual behaviour in a sample of 26 male sex offenders; 19 child sex offenders and non-child sex offenders, who had committed a contact sexual offence. Despite the use of a small sample and the absence of a control group, the study found that VT showed that the age and gender preference of the individuals were consistent with their past sexual behaviour. In addition, VT was able to differentiate individuals who had child victims from those who had adult victims. This study used both clothed and nude stimuli and it emphasised the relevance of using both types of stimuli for future research.

The use of nude stimuli is one of the main differences between the VT developed by Abel and the VT from the Gress (2001) project. The literature shows that nude stimuli provide better discrimination of an individual’s sexual interest when compared to clothed stimuli (Gress, 2005; Wright & Adams, 1994, 1999). Therefore using computer constructed images also becomes an advantage when showing nude stimuli, due to the ethical concerns described earlier.

Following the Gress (2001) project, Laws and Gress (2004) decided that it was important to develop a set of standardised computer generated images. This was due to the further unavailability of the images developed by Konopasky. The computer constructed images were developed by the Pacific Psychological Association in 2004 and took a year to be completed (Laws & Gress, 2004). The project produced two sets of 80 images each, half of the set contained nude images, while the other half included clothed images. These sets of images were to be used in two different assessments: Viewing Time (VT) and Choice Reaction Time (CRT). The CRT assessment will be explained in the next section.

Gress (2007) included the above sets of computer constructed images in her study of the VT and CRT assessments. The study found that the VT and CRT assessments produced subtest scores with high reliability, that there were significant differences between youth non-sexual offenders and a university sample when assessed with the CRT measure, but not between adult sex offenders and either youth non-sexual offenders or university students and that some of the VT subtests demonstrated good clinical utility in its ability to differentiate adult heterosexual sexual offenders from non-sexual offenders.

One of the main aspects of the Viewing time procedure is that because it is a covert measure it is less prone to faking. Imhoff et al., (2010) proposed four potential explanations to the underlying mechanisms of viewing time. The first explanation involves Deliberate Delay, which means that a person find a sexual stimulus rewarding and therefore will delay termination in order to keep viewing the image. The second explanation involves the Attentional Adhesion to sexual stimuli. In this case the response will be delayed, because the individual will be distracted by the sexual stimulus and therefore his attention is not focused on an actual task. The third potential explanation is related to the individual's hesitance to make a decision due to Sexual Content Induced Delay. The last potential explanation is related to Mate Identification. This explanation postulates when an individual looks at preferred sexual stimuli, internal processes that could be related to attention and schematic concepts are triggered. The findings of Imhoff et al., (2010) showed that none of the above accounts could clearly explain the underlying processes taken place in Viewing Time. It appeared that internal processes happen very quickly and automatically and therefore it may be harder to fake and control by participants.

In addition, clients usually believe that VT is a self-report measure and therefore the likelihood of dissimulation decreases. It has been suggested though that even if the clients were told that their viewing time was recorded, the assessment will be difficult to fake since the time is recorded in milliseconds (Williams, 2003). Relatedly the aspect of concealment of the purpose of the test has raised some concerns, which have been expressed on a website that alerts participants to the purpose of the measure and explains how VT works (InnocentDads.org, 2002). The main misconception appears to be the belief that the measure has diagnostic properties and could potentially label a

person as having a paraphilia. The article omits to explain that additional measures are used to explore a person's sexual interest and that no diagnostic category will be assigned to a client.

There is evidence in the literature to support the use of Viewing Time as a useful measure of sexual interest in children (Abel et al., 2004; Harris et al., 1996). Viewing Time has proven to be a less intrusive, easier to administer assessment and has shown good discrimination ability among individuals whose sexual interest is in children.

Choice Reaction Time

Choice Reaction Time assessments are based in the premise that individuals' reaction times will be longer when looking at a stimulus that they find sexually attractive (Laws & Gress, 2004). It is therefore possible to measure and discriminate sexual interest to various stimuli presented to a subject.

The use of Choice Reaction Time in the field of sexual interest was inspired by a theory of sexual arousal that Singer (1984) proposed. Singer pointed out the complexity of defining and conceptualizing sexual arousal. He proposed dividing sexual arousal in three components; the aesthetic response, the approach response, and the genital response. While the approach response refers to the body movements toward the desired object and the genital response refers to the somatic responses related to a sexual object, the aesthetic response involves "an hedonic feeling in response to a sexual stimulus, as for instance, the sight of an attractive face or figure or a pleasingly stimulating voice" (Singer, 1984, p.233). CRT relates to the aesthetic response. The theory follows on to explain that in the aesthetic response, the individual will try to keep the sexually attractive stimulus in view. It is assumed therefore that individuals will spend longer time looking at images of people that they find sexually attractive.

CRT was first used to assess the sexual preference of four groups; 20 heterosexual males, 20 heterosexual females, 20 homosexual males and 20 homosexual females (Wright & Adams, 1994). These groups were given the task to locate a white dot that was superimposed on 60 slides. The images included nude males, females and neutral images. The study showed that even though participants were given specific instructions

to work as quickly as they could to locate the white dot, the fact that they were looking at images of their preferred sex interfered in their ability to perform well in the task. An important finding from this study was that men appeared to spend longer time looking at the images of their preferred sexual interest and they also made more mistakes during the test than female participants (Wright & Adams, 1994)

Wright and Adams' (1994) replicated their study in 1999 (Wright & Adams, 1999). The purpose of the new study was to explore the influence of two different sets of stimuli, clothed and nude on the participants' ability to perform on a choice reaction time task. Four groups of 20 participants each, 20 heterosexual males, 20 heterosexual females, 20 homosexual males and 20 homosexual females were exposed to 60 sets of clothed, nude and neutral stimuli of both genders. The study showed that the use of both clothed and nude stimuli created interference with the CRT task. However, the use of nude stimuli created greater interference on the participants' performance on the CRT task than the clothed and neutral stimuli (Wright & Adams, 1999).

A more recent study used CRT to assess the sexual interest of 15 heterosexual men and 11 homosexual men (Santtila et al., 2009). The participants were presented with 68 pictures of males and 70 pictures of females, which were collected from different internet sites. Some of the pictures were sexually explicit (nude) and some were not (partially clothed). The results of the study indicated that participants had longer reaction times to sexually explicit stimuli of their preferred gender. Despite these results, the researchers highlighted the need to use constructed images to overcome ethical issues related to the use of images of real people.

There are two articles in the literature to date that have used the CRT assessment in a sample of adult child sex offenders. Gress (2007) used CRT and VT on 44 youth non sexual offenders, 60 university students and 22 adult sexual offenders. The participants were exposed to 40 slides that included images of people in five stages of development, ranging from prepubescent to adults. The study found that the youth non sexual offenders and adult sexual offenders had longer response times for images of females than the university students. In addition, all participants had longer response times for

images of children rather than adults. The study indicated that the results may be due VT and CRT measuring different constructs or poor design of the CRT test.

A more current study by Mokros, Dombert, Osterheider, Zappala, and Santtila, (2010) used CRT in a group of 21 child sex offenders and 21 non sex offenders. The sample for this study was gathered from high security forensic psychiatric hospitals. Participants were presented with 80 constructed images that had been developed by the Pacific Psychological Assessment Corporation in 2004 (Laws & Gress, 2004). These images included nude and clothed stimuli. Results of this study showed that participants' reaction times were longer when viewing sexually explicit images and that child sex offenders showed longer reaction times to stimuli that involved children.

It appeared that there were different outcomes in the 2007 (Gress, 2007) and 2010 (Mokros, Dombert, Osterheider, Zappala, & Santtila, 2010) studies using CRT with a sample of child sex offenders. However, it is important to consider that some of the issues surrounding Gress's (2007) study involved the small sample used, the difference in ages between the adult sample and the control group (mean age for control group was between 16.5 and 20.0, while the mean age for the adult sex offender sample was 41.1) as well as some methodological issues. Based on the findings from the four studies, as well as being mindful of the limitations of the study done by Gress (2007), the use of a CRT task to explore men's sexual interest in children appears to be important and justified. The studies support the use of both clothed and nude stimuli and the study indicates that men's sexual orientation appeared to be more accurately identified through this assessment. It has also been suggested by Gress (2007) that further studies should consider test-retest reliability on the same participants.

Purpose of this research

Rationale

Understanding the factors involved in child sexual offending is complex (Ward, 2002). Men who have offended against children pose a challenge to clinicians with regard to the amount of information they decide to disclose for treatment. As a result, emphasis is

on developing assessment measures that will provide reliable information related to the offenders' sexual interests and that will help target the offender's treatment needs. Given the limited research done on alternative assessments and the continuing legal and ethical issues surrounding phallometry, it is the intention of this research to evaluate an alternative measure that is non-intrusive and shown to be successful in the assessment of deviant sexual preference with child sex offenders in New Zealand.

This study proposes VT and CRT as potential measures that could address or reduce the invasiveness for individuals undergoing phallometric assessment. There is no current research in New Zealand on the utility of VT and CRT in the assessment of deviant sexual interests in child sex offenders. Therefore, the focus of this research will be on exploring the reliability of the VT and CRT assessments in the identification of deviant sexual interests among child sex offenders.

The New Zealand Department of Corrections provides treatment every year to child sex offenders through two Special Treatment Units, Te Piriti in Auckland and Kia Marama in Christchurch. Prior to child sex offenders undergoing treatment, they are assessed for their deviant sexual preferences. The current measure used for the assessment of deviant sexual preferences in the Special Treatment Units is the Penile Plethysmograph. Despite its use, there are issues associated with the intrusiveness of the measure, the costs associated with using the specialised equipment, training of staff, and prohibitive assessment time (Laws & Gress, 2004).

Overall, having information regarding offenders' sexual interests would be essential for their appropriate treatment and risk assessment. For that reason, the New Zealand Department of Corrections is interested in additional measures that could provide information on offenders' sexual preferences. VT and CRT are two measures that have been developed internationally and that have proven to be useful into the assessment of offenders' sexual interest (Abel et al., 2004; Harris et al., 1996).

Only a few studies have been done with a sample of adult child sex offenders, such as Gress (2005), Gress (2007) and Harris et al. (1996). This study intends to explore the utility of VT and CRT with such a sample. In addition no test-retest reliability has been

done with the measures and this study intends to fill that research gap, as recommended by Gress (2007). Following consideration from earlier studies, this research will include both nude and clothed stimuli and clients will not be told that their viewing time is being recorded, in order to strengthen the results of this study. VT and CRT assessments measure sexual preferences in a more subtle way, which could potentially address some of the issues associated with faking. If the above assessments prove to be useful, they will become important tools in both the treatment and the assessment of an offender's risk to re-offend.

Aims and Hypotheses

The aim of the present study is to explore the utility of VT and CRT in the assessment of deviant sexual interests in Child Sex offenders in New Zealand. This study will assess whether VT and CRT assessments are reliable over time and whether these assessments can discriminate child sex offenders according to their sexual deviance, the gender of their victims and the age of their victims; defined as prepubescent (12 years and under) and post pubescent (13 years and older).

The Hypotheses of this study are:

- **Hypothesis 1:** Viewing time and Choice Reaction time assessments will be reliable over time.
- **Hypothesis 2a:** Sexual deviance should influence VT and CRT results. Specifically, men who score higher on sexual deviance should have higher VT and CRT scores on more deviant images involving younger children.
- **Hypothesis 2b:** VT and CRT will reliably differentiate child sex offenders who offended only against male victims from those who offended only against female victims.
- **Hypothesis 2c:** VT and CRT will reliably differentiate child sex offenders who offended against prepubescent victims from those who offended against post pubescent victims.

This study will also examine the relationship between VT and CRT assessments as well as the relationship between responses for clothed and nude stimuli respectively.

CHAPTER TWO

METHODOLOGY

Participants

Participants in this study consisted of 52 child sex offenders who were serving a prison sentence and who were attending the initial stages of treatment at Te Piriti Special Treatment Unit (STU), Auckland Prison. As part of the entry into the treatment programme at the STU all participants had signed the form "Special Treatment Units Consent to Participate in Treatment" which is included in Appendix H. This consent form includes participant's consent for researchers to access databases and its potential use for research. Participants are given thorough explanation about the consent form before they sign it. The results of the VT and CRT assessment were not included in information collected for the purpose of psychological reports to the New Zealand Parole Board, Prison Services, and Community Probation & Psychological Services, as the study was an exploratory one.

All participants had been convicted of one or more sexual offences against children or young person under the age of 16. Data was collected over a one year period from July 2008 to July 2009. During this period fifty two child sex offenders were considered to be suitable to participate in the study. The participants showed no visual impairment, identified intellectual disability, motor impairment or illiteracy in English that prevented them from completing the VT and CRT assessments.

Demographic information on participants for this study was gathered from Te Piriti STU's database. Table 1 outlines the demographic characteristics of the sample. As can be seen, participants' ages ranged between 21 and 81 years old. Most identified as New Zealand Maori. The participants' offences ranged from indecent assault to rape. There were eight participants who had male victims only, 36 had female victims only and eight had both male and female victims. Participants included in the study involved those who were given either a sentence end date or an indefinite sentence of Preventive Detention.

The sample for this study was not representative of the general child sex offender population who attends treatment for their sexual offending. David T. Jones, Senior Clinical Psychologist at Te Piriti STU, reported the demographics of child sex offenders who had attended treatment from 1999 to 2007 at Te Piriti or Kia Marama. It was found that European offenders who attend treatment make up 64% of the total of treated offenders. In addition, child sex offenders with victims age 12 years and younger made up for only 25% of offenders who attended treatment (D. T. Jones, personal communication, September 15, 2011). In this study, offenders who were of Maori descent (53.8%) and offenders whose victims were 12 years old and younger (63.5%) were overrepresented in this study.

Table 1

Demographic Characteristics of the Sample.

Demographics (N=52)	Frequency (n)	Percent
Ethnicity		
Maori	28	53.8
European/Pakeha	21	40.4
Pacific Islander	3	5.8
Age (*mean = 45.73, SD = 12.94)*		
21 - 30	7	13.5
31 - 40	13	25.0
41 - 50	16	30.8
51 - 60	9	17.3
61 - 70	5	9.6
71+	2	3.8
Victim gender		
Female only	36	69.2
Male only	8	15.4
Both	8	15.4
Victim age		
12 and under, only	33	63.5
13 and over, only	7	13.5
Both	12	23.1

Design of the Research

Approval for this study was obtained from the Massey University Human Ethics Committee and the Department of Corrections to collect and use the data for research. This study used a test-retest design. Participants were assessed with the VT and CRT assessments at two points in time. This was to explore the test-retest reliability of the measures.

Research Setting

Te Piriti Special Treatment Unit (STU) is one of two prison based units that have been created for the exclusive treatment of child sexual offenders in New Zealand. Te Piriti STU is the sister unit of Kia Marama STU and is based at Auckland prison, while Kia Marama is located in Christchurch. Te Piriti STU was established in 1994 due to the positive results experienced at Kia Marama and the increased need to provide treatment to sexual offenders from the North Island. Te Piriti STU is a low/medium security, stand-alone facility and houses 60 prisoners at one time. The therapy building is separate from the prisoner's compound but still within the prison compound. It is in this building that the therapy programmes are run.

Attendance to the programme is on a voluntary basis. However, to be accepted into the programme, prisoners' risk rating should be minimum to low medium, they should have acknowledged that they have committed a sexual offence, that they consent to treatment and they should not present with any serious neurological damage. In addition, it is expected that prisoners complete the treatment programme by the two thirds' date of their sentence. Prisoners are usually transferred to the unit at least one year prior to their two thirds' date to begin treatment.

Te Piriti STU has a bicultural approach to treatment and this is what distinguishes Te Piriti from Kia Marama and other programmes that are run in other countries (Larson, Robertson, Hillman, & Hudson, 1998). Te Piriti STU has integrated the Bicultural Therapy Model into the treatment of Maori offenders. The purpose of this is to assist offenders in their wellbeing, help them in their rehabilitation process as well as creating an environment that is culturally safe. The cultural component of the programme aims at

helping Maori offenders connect to their cultural identity and to motivate them into adopting a more positive lifestyle.

The treatment provided at Te Piriti STU has a Cognitive Behavioural Therapy (CBT) focus. In order to connect CBT concepts with Maori concepts, the Cultural Consultant of the unit attends some of the therapy sessions and provides a thorough explanation of the Maori Tapa Wha model, which is the equivalent of the CBT focus. By doing this, Maori offenders can better understand and relate to Western approaches and therefore embrace the CBT concepts. Values and protocols from Maori culture are also integrated into the unit in general, by means of Karakia (prayer), Waiata (song), Pōwhiri (formal welcome), Whakatau (informal welcome) and Hākari (feast and meal). Cultural supervision is provided for both staff and inmates. Additional cultural activities are available for prisoners such as Kapa Haka (performing arts), Art, Weaving, Te Reo lessons and Carving.

Te Piriti runs two treatment programmes; an adapted programme and a rolling programme. The adapted programme is for prisoners who have low intellectual functioning. The rolling programme is an open programme that consists of three stages. The purpose of the first stage is to help prisoners get orientated and settle into the unit, become familiar with group work, giving and receiving feedback as well as receiving explanation of the treatment programme. This initial stage also involves an assessment phase. Prisoners initially attend a two hour session once or twice a week. This stage of treatment lasts approximately between six to ten weeks. During the second stage, which is the core programme, prisoners attend three hours session three times a week. Prisoners often stay in this stage for about six months, but they can stay longer if this suits their treatment needs. In the core stage, prisoners focus on their sexual offending and explore the aspects that led them to commit their crimes. Once prisoners have identified the contributing factors to their offending, and have learned the skills to manage them, they are interviewed by an exit panel. The exit panel includes the Principal Psychologist, The Principal Corrections Officer and the prisoner's therapist. This panel will determine whether the prisoner is ready to leave this phase of treatment and move onto the third and final stage, which is Maintenance. Prisoners usually stay in this stage of treatment for about nine weeks. In this last phase prisoners are expected to

put into practice the skills learned throughout the earlier phases of treatment and to be able to manage their identified issues. Prisoners often stay in this phase until they are released from prison. Overall the programme takes about 40 weeks.

Principles of a Therapeutic Community were incorporated into the treatment at Te Piriti STU in 2004. The term community involves the active involvement of staff and prisoners into the therapeutic process. The purpose of the therapeutic community is to create a therapeutic environment in itself, where prisoners have the opportunity to practice new skills, learn to interact and respect each other and to take responsibility for their behaviours. Prisoners benefit from this approach since they can receive feedback about their unhelpful interactions in many different contexts and not just the therapy room. Te Piriti STU provides an environment where prisoners can feel encouraged to change and to carry this change back into the community once they are released from prison.

Measures

Participants are assessed with the STABLE-2007 as part of standard assessment at Te Piriti STU. Te Piriti STU installed the VT and CRT software in the unit's computer on 1 July 2008. These assessments became standard assessments in the unit as part of a trial to assess the validity and reliability of the VT and CRT assessments. The data collected from these assessments was then added to the unit's research database.

The VT and CRT assessments are the same measures that were used in Gress' (2007) study. The time spent by the individual viewing each stimulus is the dependent variable in both instances, and the independent variables in this study are sexual deviance (as measured by the STABLE-2007), victim gender (male, female, or both) and victim age (pre-pubescent, post pubescent or both). The difference in time is measured in milliseconds, and is unlikely to be under the control of the client. This means that it is less likely to be faked. The assessment should take approximately 1 hour to complete. Participants of this study took approximately 1 hour and 45 minutes on the VT and CRT assessments in total. The actual VT and CRT assessment will be described below. Scale reliability for the measures will be reported in the Results section.

Viewing Time

The VT assessment consisted of 10 practice items and 415 test items. The test items were constructed images of people developed by Laws and Gress (2004). The images represent the five different stages of physical development as defined by the Tanner Sexual Maturity Scales (Tanner, 1962). These stages of physical development range from prepubescent to adults. The scale uses external primary and secondary sex characteristics for the measurement of physical development on individuals, such as the pubic hair and size of genitals for boys and breast development and pubic hair for girls. Because every person is different, people pass through the stages at different times (Tanner, 1962). Fuller, Barnard, Robbins, and Spears (1988) considered that the use of the Tanner scales will prevent the use of age categories in an arbitrary way and will allow for standardisation of the stimuli used in the assessment of sexual interest.
For both genders there were four clothed images and four unclothed images per Tanner category and each image appeared five times for a total of 40 items per Tanner category and 200 for each gender. Fifteen neutral items were included in the assessment and these were randomly placed in between the other items. The neutral items were five images of landscapes that Gress (2007) had downloaded from the internet, each repeated three times with a different neutral question such as "Would you like to visit this place?"

The pictures used in the VT are regarded as non-erotic, in that the subjects are standing still in natural, non-provocative poses, with only one subject visible at any one time (Laws & Gress, 2004). The picture is located in the middle of the screen and a question related to the image is placed above the image for the participant to answer. There were five questions and each question was paired with each image in a way that it balanced the influence of the question on the image, such as the time taken to read the question (Gress, 2007).

In addition the images were presented in a previously determined random order. Participants answered the questions by choosing one of five Likert type answers located on the right hand side of the picture. Participants were asked to press a number (1 to 5) on the keyboard that would indicate the answer that they chose. Once the participant had chosen his answer a new slide will appear on the screen.

In the output from the programme, which is used for analysis, each item is given a label, such as ‘FT1C-3-A’, consisting of the following:

1. Gender (F or M), or Neutral for the neutral items.
2. Tanner Physical Development Category (T1 to T5); not applicable for neutral items.
3. A ‘C’ is then appended if the item represents a clothed image; not applicable for neutral items
4. Image number (1 to 4) within the Tanner category.
5. Character to indicate which question was given:
   A = “How sexually attractive is this person?”
   F = “Would you like to be friends with this person?”
   O = “How old is this person?”
   P = “How pretty/handsome is this person (regardless of age)?”
   S = “Would you like to have sex with this person?”

Choice Reaction Time

The CRT measure consisted of 10 practice items and 425 test items. The images are identical to those used in the VT assessment, but instead of participants having to answer a question they must instead locate a white dot superimposed on the image in one of five different locations (top right, top left, centre, bottom right and bottom left). For both genders there were 4 clothed images and 4 unclothed images per Tanner category and each image appeared five times as per the five dot locations, for a total of 40 items per Tanner category and 200 for each gender. Twenty five neutral items were included in the assessment and these were randomly placed in between the other items. These were the same five neutral images that were also used in the VT assessment and these were presented five times each; one for each of the five possible dot locations.

Participants had to decide where the dot was located and choose their answer from the options located on the right hand side of the image. As in the VT assessment, once the

participant had chosen his answer a new slide will appear. In addition, participants were told to answer the questions as fast and as accurately as possible.

In the output from the program, which is used for analysis, each item is given a label, such as 'FT1C-2-BL', consisting of the following:

1. Gender (F or M), or Neutral for the neutral items.
2. Tanner Physical Development Category (T1 to T5); not applicable for neutral items.
3. A 'C' is then appended if the item represents a clothed image; not applicable for neutral items.
4. Image number (1 to 4) within the Tanner category.
5. Code to indicate the location of the white dot:
   TL = Top left
   TR = Top right
   M = Middle
   BL = Bottom left
   BR = Bottom right

STABLE-2007

The increased need to identify useful dynamic risk factors for the monitoring and management of sex offenders led Hanson and Harris (1998, 2000) to conduct the Dynamic Predictors Project in Canada. Dynamic risk factors have been found to correlate with risk of re-offending and they are factors that are amenable to change over time (Harris & Hanson, 2010). Hanson and Harris' (1998, 2000) project involved gathering information from follow up studies, as well as extensive file reviews and interviews with supervision officers. Their sample involved a group of non-recidivists (201) and recidivists (208) sex offenders that had offended while on community supervision. For the recidivists' group the data was collected at two points in time, six months and one month prior to re-offending. The purpose of this was to identify "stable enduring characteristics of the offender, which we called stable dynamic risk factors, from acute risk factors, which are temporarily transient environmental and mood states"

(Harris & Hanson, 2010, p. 301). Following this project the Sex Offender Need Assessment Rating (SONAR) was designed in order to assess stable and acute changes in sexual offenders. However there were some issues with the measure, such as the study being retrospective, scoring procedures and the exclusion of some important criminogenic variables.

Due to issues with the SONAR, two measures were developed in 2000, the STABLE 2000 and ACUTE 2000 (Harris & Hanson, 2010). The rationale for the above division was that stable factors helped in the identification of treatment needs, while acute factors helped in the management of high risk situations while the offender was on supervision. The STABLE 2000 consisted of sixteen items divided in six subsections. In order to validate these measures empirically, Hanson and Harris decided to conduct the Dynamic Supervision Project (Hanson, Harris, Scott & Helmus, 2007). This was a prospective study that involved the training of community supervision officers on three measures, the STATIC-99, the STABLE 2000 and the ACUTE 2000. All provinces and territories of Canada were involved in the project as well as the States of Alaska and Iowa. The sample included 997 offenders who were starting a period of community supervision for a sexual offence. The STABLE-2007 and ACUTE 2007 were the outcome of the Dynamic Predictors Project (Hanson & Harris, 1998). The ACUTE 2007 is primarily used with offenders undergoing supervision in the community and the STABLE-2007 is used with offenders serving a prison sentence.

This research will focus on child sex offenders serving a prison sentence. There is no current research on the inter-rater reliability of the STABLE-2007 due to this being a new revised measure of the STABLE-2000. However, Hanson and Harris (2007) evaluated the reliability of the STABLE-2000 via two methods. The first method involved comparing the responses of 213 officers trained by the researchers and 45 officers trained by three other trainers. Hanson and Harris (2007) found that between 47% and 67% of the officers trained by them were within one point of the correct answer and between 74% and 99% were within three points of the correct answer. For officers trained by other trainers 28% and 80% of them were within one point of the correct answer and between 80% and 100% were within three points of the correct answer (Hanson & Harris, 2007).

The second method involved the file review of 92 cases. The intraclass correlation for the STABLE 2000 was .89 for total scores. The results of this second method indicated more the understanding officers had of the coding rules rather than the concordance of their individual assessments (Hanson & Harris, 2007). The STABLE-2007 contains 13 items divided into 5 sections. Three items were dropped from the STABLE-2000, since it was found that they were not associated with sexual recidivism. The STABLE-2007 sections are Significant Social Influences, Intimacy Deficits, General Self-regulation, Sexual Self-regulation and Cooperation with Supervision.

The Section of Sexual Self-regulation contains the item Deviant Sexual Preference. The score on this item is based on four domains. These are the number of sex offence victims, the number of deviant preference victims/activities, self-report of deviant history or preferences and results of specialized testing (phallometric assessment). The lowest score that can be obtained is zero, which indicates that there is no evidence of deviant sexual preferences, a score of one will indicate that there is some evidence that the offender has some deviant sexual preferences, and a score of two will indicate that there is evidence that the offender has deviant sexual preferences.

One of the advantages of the STABLE-2007 is that it allows observation of change on the level of risk of an offender over a period of time. In addition, it is user friendly and it can be administered by different professionals dealing with sex offenders in different settings, which increases the likelihood of earlier detection and prevention of further offences (Harris & Hanson, 2010). Due to this measure being available to different professionals and being flexible with the skills needed to administer it, the STABLE-2007 has become one of the most used instruments in the assessment of dynamic risk among sex offenders in countries such as Canada and United States. In addition, several other countries have also received training in the measure, including New Zealand (Harris & Hanson, 2010).

The New Zealand Department of Corrections first decided to adopt the STABLE-2000 measure in 2006. The Te Piriti and Kia Marama STUs currently use the STABLE-2007 in the assessment of dynamic risk among child sex offenders. This measure is used in

conjunction with actuarial measures of risk in order to provide an overall risk estimate of sexual recidivism. The STABLE-2007 has become a standard assessment for every child sex offender attending treatment at Te Piriti Special Treatment Unit. Offenders attending treatment are usually assessed with the STABLE-2007 before attending the core treatment programme and once they complete treatment their scores on the measure are reviewed and at times adjusted, to assess post treatment effects. This study will use the scores obtained from child sex offenders attending the Te Piriti STU on the item Deviant Sexual Preferences pre-treatment to correlate it with the scores on the VT and CRT assessments to assess their convergent validity.

Procedure

Participants for this study were child sex offenders attending the early stages of treatment at Te Piriti STU. Participants were assessed with the VT and CRT at two points in time (Test 1 and Test 2). Both tests took place while offenders were attending the first stage of treatment. Test 1 took place a couple of weeks after participants had arrived in the unit. There are no specific timeframes with regards to when Test 2 should be administered (Hogan, 2003; Kaplan & Saccuzzo, 2005; Kline, 2005). Test 2 took place three to four weeks after Test 1 and before the participants moved on to the core treatment stage.

When doing the test, each participant was taken to the therapy room were the assessment was conducted. This therapy room is mainly used for the Plethysmograph assessment and the environment is set up for that purpose. The room is divided in two and access to each half is through a door. The first half is for the assessor to sit in and the other one is for the individual when being assessed with the Plethysmograph. The location of this therapy room provided a quiet environment and allowed each participant to focus on the assessment without any distraction.

Before Test 1 was conducted, participants were given an overview of the VT and CRT assessments and were explained that these assessments measured sexual interest. Participants were informed that this assessment involved two parts and that they could have a break in between them if they needed it. Each participant was explained that the

measures included constructed images of people from both genders and that some images were clothed and some were unclothed. Participants were also informed that the assessment will take approximately one hour to complete. The aims and the nature of the assessments were not disclosed to participants. In addition, participants were given the opportunity to ask any further questions.

Followed by this explanation, participants were taken to the room where the assessment was conducted. They were seated in front of the lap top computer and they were asked to read the instructions page for the VT assessment. Any questions from the instructions page were answered. Each participant was shown how to operate the keyboard and explained that to answer the questions they had to select a number from 1 to 5. These numbers were located to the left on the upper part of the keyboard.

Before the real VT assessment was presented, each participant first had to complete a practice VT assessment. The researcher left the assessment room and waited in the room next door until each participant completed the practice test. Once the participant made the researcher aware that he had completed the practice test, the researcher returned to the assessment room and answered any further questions. Thereafter, the researcher returned to the room next door and the participant continued with the real VT assessment. On completion of the VT assessment, the participant notified the researcher by calling out, upon which the researcher came back to switch to the CRT. Hereafter the participant proceeded to read the CRT's instruction page, completing the CRT practice test and finally the real CRT test.

After the participant had completed both the VT and CRT assessments, the participant returned to the researcher's room and was asked for comments on the measures. Each participant's comments were recorded. Each participant was informed that Test 2 will take place in about three to four weeks' time. For Test 2, participants were informed that they will be completing the same assessment they completed three to four weeks' ago. They were given the opportunity to ask any questions and thereafter they were conducted to the assessment room.

The results of each participant were saved in the laptop computer under an ID number randomly generated by the VT/CRT software. The assessment data was recorded as a data file on that computer, and was also backed up on a DVD-ROM. The electronic files were password protected and these files were not available in any other computer at Te Piriti STU. In order to ensure anonymity, participants' names were not stored on the computer. Once the ID number was created for each participant, both the ID number and the name of the participant were given to the Executive Officer of Te Piriti STU. The Executive Officer used this information to collect each participant's STABLE scores and demographic information from Te Piriti STU's database and the researcher received this information under ID numbers.

Data analysis of the participants' results started in 2011. The reason for this was to minimise any researcher bias, since the researcher had a therapeutic relationship with each participant at the start of their treatment. It was expected that by 2011 all participants had completed the treatment programme and were likely to have left Te Piriti STU.

Ethical Considerations

Consent to participate in research should be given by a person who is fully informed, who decides to give consent voluntarily and who shows ability to make that decision (Eyler & Jeste, 2006). Offenders' participation in research studies has raised issues with regards to their ability to give consent voluntarily (Elger 2008). Given this, offenders at Te Piriti STU are given a thorough explanation of the "Special Treatment Units Consent to participate in Treatment" as soon as they arrive to treatment. The form is read to them and they are given the opportunity to take the form with them to read before they sign it.

The consent form states that information gathered in the unit about the prisoners through assessment will be entered in Te Piriti's database. Participants are made aware that this information has the potential to be used for research purposes. It is explained to offenders that if their information was to be used their anonymity will be ensured. This study ensured offender's anonymity by using ID numbers when gathering the data, and that the data would not be analysed until they had completed their treatment in the STU.

Data Analysis

The data collected for this study was entered into the Statistical Package for Social Sciences Software, Version 17 (SPSS-17, 2011), which was used to screen and analyse the data. Data screening involved the adjustment of univariate outliers for VT and CRT (for both Test 1 and Test 2) with Winsorising (Miller, 1991). This will be explained in detail in Chapter 3.

The Winsorised VT and CRT data from Test 1 were then collapsed from 415 and 420 categories respectively to 11 categories each. Descriptive statistics were done on the Winsorised and reduced data for VT and CRT at Test 1 only. The normality of the distribution of scores was evaluated using histograms and the Shapiro-Wilk test. Furthermore, the data was analysed for scale reliability.

Power Analysis

A power analysis was conducted in order to determine the ability of the analyses to identify present effects. Table 2 shows the required minimum sample sizes for those parametric analyses that would be relevant for this study, in order to obtain a power of .80 for a medium effect size (Cohen, 1992). Since there are less than 64 cases available for the test-retest analysis, it is evident that the alpha level would have to be adjusted to .10 to use a t-test as part of this analysis. Such an adjustment increases the risk of a Type One error. Also, the analysis required for testing Hypothesis 2a, 2b and 2c will involve comparing groups of cases within the same sample (51 for VT and 52 for CRT) and will therefore have fewer than the required minimum of 50 cases per group for 2 groups, or 41 cases per group for 3 groups. Therefore for this study the non-parametric Wilcoxon signed Ranks test is used instead of t-test for the test-retest analysis, and the non-parametric Kruskal-Wallis test is used instead of ANOVA, since these are not similarly dependent on sample size. Monte Carlo simulations have also shown the Kruskall-Wallis test to have higher statistical power than the traditional ANOVA when the data is not normal distributed (Van Hecke, 2010), which is typically the case with latency data.

Table 2

Required Sample Size for Relevant Analyses.

Analysis	$\alpha = .05$	$\alpha = .10$
t-test	64	50
ANOVA, 2 groups	64	50
ANOVA, 3 groups	52	41

CHAPTER THREE

RESULTS

Data Screening

The VT data for the Test 1 was adjusted for univariate outliers with Winsorising, by replacing all latencies less than 300ms with 300ms and all latencies longer than three standard deviations above the mean with the latency corresponding to three standard deviations above the mean, as recommended by Miller (1991). This was computed separately for each participant. The cut-off of three standard deviations was chosen to reduce bias for the means to be calculated (Miller, 1991). This resulted in only relatively few changes to any participants' raw values, with an average of 1.77% of participants' raw response times being changed and the highest and lowest percentage of changed response times for a single participant being 3.13% and 0.24% respectively. Based on this all cases were included in the further analysis. Gress (2007) suggests dropping the case from analysis if the adjustment changes the raw response scores in more than 5% of a participant's raw scores.

One case had very significant outliers in the VT test, all ranging from 55 to 153 seconds. This contrasted heavily with outliers for other cases which were typically 10s of seconds. This case also had the highest standard deviation at 14.58 seconds. It was noted in the comments to the VT test that this particular participant thought the test was confusing and was also talking with the test administrator during the testing. It is therefore assumed that none of these extreme outliers represent real response times, and the case was dropped from the VT analysis, thereby reducing the sample size for VT from 52 to 51.

Similar to the VT data, the CRT data for Test 1 was also Winsorised by replacing all latencies less than 300ms with 300ms and all latencies longer than three standard deviations above the mean with the latency corresponding to three standard deviations above the mean. This was again computed separately for each participant. An average of 1.97% of participants' raw response times was changed and the highest and lowest

percentage of changed response times for a single participant was 4.24% and 0.24% respectively. Based on this all cases were included in the further analysis.

Some interesting results were however found when looking at the distribution of outliers across the items. In total 34 out of 425 items (8%) had more than 5% of participants' response times adjusted. To identify underlying causes for this it should be considered that in general the CRT data differs significantly in nature from the VT data because reaction time measurements result in shorter viewing times. If the outliers among the CRT response times were genuine reflections of the participants' actual interest in the images themselves then one would expect to see similar long response times for all, or most of, the identical test items (i.e., those where only the location of the white dot differs). However, some particular single test items showed a very significant amount of outliers, with a maximum of 67% of participants' response times adjusted for one particular item; FT5-1-BR. Table 3 shows the distribution of outliers for the FT5-1 items. All five test items are identical images, except for the location of the white dot.

Table 3
Distribution of Outliers for the Five CRT FT5-1 Items.

Item	FT5-1-BL	FT5-1-BR	FT5-1-M	FT5-1-TL	FT5-1-TR
Outliers	**13.46%**	**67.31%**	0.00%	1.92%	0.00%

Upon inspection of the particular test images for these items it was obvious that there were problems with the test material. For item FT5-1-BR it was discovered that the white dot was located on a very light part of the image where it was practically invisible compared to the other FT5-1 images. Also for item FT5-1-BL a similar situation occurred, although it was slightly easier to identify the dot in this case.
The image for each test item with more than 5% outliers was then inspected to identify if similar problems might account for the outliers.

Table 4 shows the findings, including the already described cases of FT5-1-BR and FT5-1-BL.

Table 4

CRT Items with Large Numbers of Outliers.

Item	Outliers	Observation
FT5-1-BR	67.31%	Dot on light skin, nearly invisible
FT5-1-BL	13.46%	Dot on hand, nearly invisible
FT5C-1-BR	15.38%	Dot on light skin, nearly invisible
FT5C-1-BL	15.38%	Dot on hand, nearly invisible
FT4-1-M	9.62%	Dot on light skin, nearly invisible
MT5-3-M	11.54%	Dot on light skin, nearly invisible
Neutral1-TR	65.38%	First image in the test

The test item Neutral1-TR had 65% of participants' response times identified as outliers. The test images are always shown in the same order and it appears that Neutral 1-TR is the first image being shown. It may have been that participants were still becoming familiar with the identification of the dot at this stage. Several participants also mentioned that some of the dots in the CRT test were difficult to locate. This seems to be confirmed by the findings above.

Overall it seems reasonable to believe that the outliers for the items in Table 4 are due to the participants having difficulty in locating the white dot. As a result of this the items FT5-1-BR, FT5-1-BL, FT5C-1-BR, FT5C-1-BL, FT4-1-M, MT5-3-M and Neutral1-TR was excluded from further analysis. No other obvious reasons could be identified for the remaining of the 8% of items with more than 5% of participants' response times adjusted. Therefore no further items were excluded since these may still represent genuine response times.

The VT and CRT data for the retest also contained outliers as identified during Winsorising.

For the VT Test 2 an average of 1.77% of participants' raw response times were identified as outliers and the highest and lowest percentage of outliers for a single participant being 4.58% and 0.48% respectively. Based on this all cases were included in the further analysis.

For the CRT Test 2 an average of 2.06% of participants' raw response times were identified as outliers and the highest and lowest percentage of outliers for a single participant being 7.76% and 0.71% respectively. The single case with 7.76% outliers only had a maximum viewing time of 14.34 seconds, which was considered to be well within the expected range. There was no reason to believe that the outliers for this case were not genuine viewing times and the case was therefore included in the analysis.

Some of the problematic items that were listed in Table 4 also had many outliers in the CRT Test 2. However, for 4 out of 7 of these the percentage of outliers was reduced to the point where it was no longer significant. Three of these items still had a very high percentage of outliers though, as illustrated in Table 5, and since these represent the same images as in the Test 1 the cause for the outliers were assumed to be the same again.

Table 5

CRT Test 2 Items with Large Numbers of Outliers.

Item	Outliers	Observation
FT5-1-BR	69.23%	Dot on light skin, nearly invisible
FT5-1-BL	30.77%	Dot on hand, nearly invisible
Neutral1-TR	78.85%	First image in the test

As a result of this the items FT5-1-BR, FT5-1-BL and Neutral1-TR were excluded from further analysis for the CRT Test 2.

Data Aggregation

Correlation between Clothed and Unclothed Stimuli

For the analyses that follow it was necessary to at least aggregate the data into the five Tanner categories. It would be natural to also make a distinction between the clothed and unclothed (nude) stimuli. Since VT and CRT must be considered a measure of sexual interest, it would be expected that response times for unclothed stimuli would be higher than for clothed, in line with the literature (Gress, 2005; Mokros et al., 2010; Wright & Adams, 1994, 1999). It would seem reasonable that participants would be more sexually interested in unclothed stimuli than in clothed stimuli.

To analyse the relationship between response times for clothed and unclothed stimuli, the data was collapsed into two categories of simply clothed and unclothed. Scatterplots were generated for these categories, for both VT and CRT separately. These are included below as Figure 1 and Figure 2. From the distribution of values on the scatterplots it was clear that response times for clothed and unclothed stimuli were highly correlated for both VT and CRT.

Figure 1

Scatterplot for Clothed and Unclothed VT Response Times.

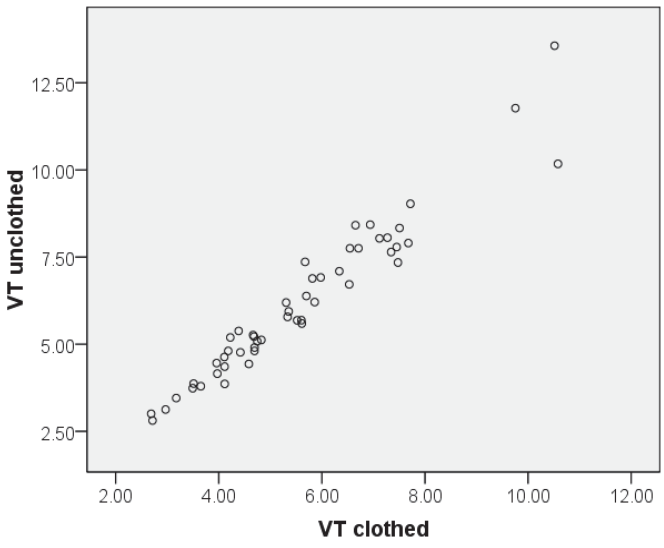
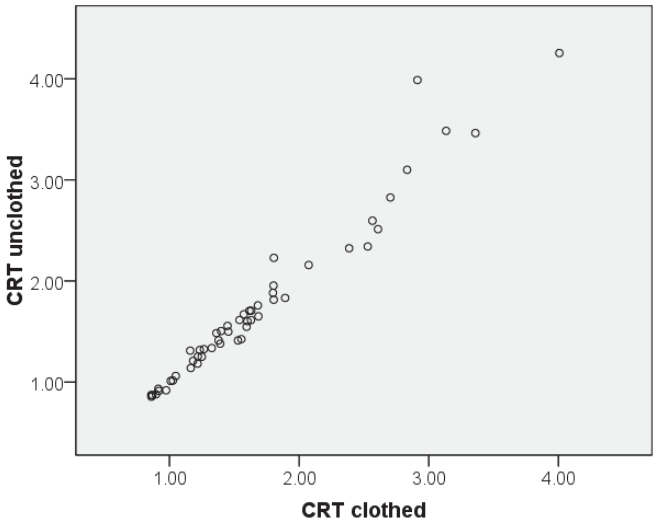


Figure 2

Scatterplot for Clothed and Unclothed CRT Response Times.



The relationship between clothed and unclothed stimuli for both VT and CRT was further investigated using Pearson product-moment correlation coefficient. As expected, based on the examination of the scatterplots, statistically significant and strong positive correlations were found between clothed and unclothed stimuli response times for both VT and CRT. The correlation coefficients are listed in Table 6 below.

Table 6

Pearson Product-Moment Correlations between Clothed and Unclothed Response Times.

	Pearson Correlation	Sig. (2-tailed)
VT (N=51)	.967**	.000
CRT (N=52)	.979**	.000

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

So, surprisingly, it appears that it makes no difference to the response times whether the presented stimuli was clothed or not. This is unexpected and will be interpreted in the Discussion later.

Since there seems to be no distinction between the responses for clothed and unclothed stimuli it was decided to collapse the Winsorised VT and CRT data into the female and

male Tanner categories, plus one for the neutral images. This resulted in the data being reduced from 415 VT items to 11 VT items, and from 418 CRT items to also 11 CRT items. Each collapsed item thereby representing the mean of the response times for all images in a particular tanner group, both clothed and unclothed. For example, all images from the Female Tanner 1 category were combined to a new item named FT1, and likewise all images from the Male Tanner 1 category were combined into MT1. Table 7 shows the combined items that will be used in the rest of the analyses.

Table 7

Combined items used in the analyses.

Name	Description
FT1	All images from Female Tanner 1 category
FT2	All images from Female Tanner 2 category
FT3	All images from Female Tanner 3 category
FT4	All images from Female Tanner 4 category
FT5	All images from Female Tanner 5 category
MT1	All images from Male Tanner 1 category
MT2	All images from Male Tanner 2 category
MT3	All images from Male Tanner 3 category
MT4	All images from Male Tanner 4 category
MT5	All images from Male Tanner 5 category
Neutral	All images from Neutral category

Descriptive Statistics

The data was inspected to evaluate whether the underlying statistical assumptions were met for the following analyses, which were Pearson Product-Moment Correlations, Wilcoxon Signed Ranks tests and Kruskal-Wallis tests.

Descriptive statistics was performed on the Winsorised and reduced data, for Test 1 only. The results are displayed in Table 8 and Table 9. For VT individual means ranged from 5.295 to 6.347 and standard deviations ranged from 1.923 to 2.251. For CRT the individual means ranged from 1.562 to 1.819 and standard deviations ranged from .663 to .953. Overall the values are positively skewed as may be expected for latency scores (Miller, 1991).

Table 8

Descriptive Statistics for VT Categories.

N=51	Min (seconds)	Max (seconds)	Mean (seconds)	SD	Variance	Skewness	Kurtosis
FT1	2.571	12.840	5.641	2.07	4.317	1.143	1.736
FT2	3.066	14.821	6.227	2.25	5.065	1.376	3.399
FT3	3.064	13.874	6.344	2.20	4.852	1.029	2.025
FT4	2.884	11.517	6.347	2.06	4.245	.609	.195
FT5	2.605	10.944	5.901	2.00	4.003	.565	-.265
MT1	2.584	11.298	5.295	1.92	3.697	1.026	1.140
MT2	2.475	11.889	5.627	2.13	4.538	1.108	1.248
MT3	2.672	12.997	5.920	2.19	4.835	1.158	1.831
MT4	2.620	12.026	6.026	2.03	4.155	.791	.642
MT5	2.595	11.491	5.575	2.04	4.199	1.036	1.067
Neutral	2.183	10.411	5.903	1.94	3.779	.199	-.689

Table 9

Descriptive Statistics for CRT Categories.

N=52	Min (seconds)	Max (seconds)	Mean (seconds)	SD	Variance	Skewness	Kurtosis
FT1	.866	3.945	1.665	.690	.476	1.278	1.550
FT2	.796	3.709	1.706	.740	.547	1.121	.532
FT3	.796	4.157	1.562	.737	.543	1.595	2.545
FT4	.792	3.828	1.698	.752	.566	1.147	.678
FT5	.857	6.174	1.819	.953	.907	2.389	7.886
MT1	.868	4.463	1.812	.839	.705	1.475	1.943
MT2	.814	3.965	1.592	.663	.440	1.658	3.154
MT3	.799	4.564	1.691	.850	.722	1.776	3.087
MT4	.858	5.418	1.745	.857	.735	2.218	6.289
MT5	.798	4.154	1.649	.708	.501	1.582	2.978
Neutral	.913	4.333	1.786	.733	.537	1.477	2.357

Tests of normality were also performed on the reduced data. The Kolmogorov-Smirnov test is traditionally used but is more accurate for large sample sizes. For the current study, the Shapiro-Wilk values seem to provide better agreement with the profiles of the histograms in Appendix A. The results for both Kolmogorov-Smirnov and Shapiro-Wilk tests of normality are displayed in Table 10 and Table 11 for VT and CRT categories respectively.

Table 10

Tests of Normality for VT Categories.

N=51	Kolmogorov-Smirnov		Shapiro-Wilk	
	Statistic	Sig.	Statistic	Sig.
FT1	.127	.038	.924	.003
FT2	.100	.200*	.907	.001
FT3	.088	.200*	.929	.004
FT4	.095	.200*	.963	.108
FT5	.144	.010	.959	.078
MT1	.118	.073	.926	.003
MT2	.120	.064	.921	.002
MT3	.110	.177	.922	.003
MT4	.137	.018	.957	.065
MT5	.163	.002	.915	.001
Neutral	.063	.200*	.983	.673

*. This is a lower bound of the true significance.

Table 11

Tests of Normality for CRT Categories.

N=52	Kolmogorov-Smirnov		Shapiro-Wilk	
	Statistic	Sig.	Statistic	Sig.
FT1	.176	.000	.886	.000
FT2	.202	.000	.880	.000
FT3	.203	.000	.838	.000
FT4	.186	.000	.883	.000
FT5	.209	.000	.776	.000
MT1	.173	.000	.854	.000
MT2	.197	.000	.846	.000
MT3	.195	.000	.809	.000
MT4	.238	.000	.782	.000
MT5	.174	.000	.864	.000
Neutral	.169	.001	.869	.000

According to the Shapiro-Wilk results only the VT categories FT4, FT5, MT4 and Neutral show sig values above .05 and can then be assumed to have a reasonably normal distribution. For CRT none of the categories have sig values above .05 and this also fits with the CRT values being more positively skewed than the VT values. The same is evident when looking at the histograms in Appendix A. It is important to realise that for reaction time data the underlying distribution is unknown, and therefore it cannot be assumed that values should be normally distributed (Whelan, 2008). The data was not transformed in order to avoid a loss of information, but this may then potentially decrease the power of the analyses.

Scale Reliability

A reliability analysis was done for both VT and CRT data for all the reduced categories, including the neutral category.

Cronbach's Alpha for the 11 categories was .986 and .990 for VT and CRT respectively, which indicated very high internal consistency. This may indicate that each category measures the same underlying factor. This factor may be different for VT and CRT though, as suggested by Gress (2007).

For both VT and CRT further reliability analyses were also done for the 5 female and male Tanner categories separately. All showed similar high internal consistency, with Cronbach's Alpha being .975 for VT female and .986 for VT male, .982 for CRT female and .984 for CRT male. The reliability statistics are shown in Table 12 below.

Table 12

Reliability Statistics.

	Cronbach's Alpha	N of Items
VT		
All, incl. Neutral	.986	11
Female only	.975	5
Male only	.986	5
CRT		
All, incl. Neutral	.989	11
Female only	.982	5
Male only	.984	5

Inter-Correlations

Before continuing with the hypothesis testing, the relationship within VT categories and within CRT categories as well as between VT and CRT categories was investigated using Pearson product-moment correlation coefficient. The complete inter-correlation matrices for VT, CRT and VT/CRT are included in Appendix B.

Statistically significant correlations were found between almost all of the paired categories. The smallest significant correlation, $r = .279$, $n = 51$, $p < .047$, was observed between VT MT1 and CRT FT5 and the largest significant correlation, $r = .564$, $n = 51$, $p < .0005$ was observed between VT MT4 and CRT MT3. All correlations were positive, with longer VT response times corresponding to longer CRT response times. Similarly, the relationship within VT and CRT categories was also investigated using Pearson product-moment correlation coefficient. Statistically significant strong positive correlations were found between all of the paired categories. Within VT the smallest significant correlation, $r = .710$, $n = 51$, $p < .0005$, was observed between VT Neutral and VT FT5 and the largest significant correlation, $r = .974$, $n = 51$, $p < .0005$ was observed between VT MT3 and VT MT2. Within CRT the smallest significant correlation, $r = .711$, $n = 52$, $p < .0005$, was observed between CRT MT5 and CRT FT5 and the largest significant correlation, $r = .976$, $n = 52$, $p < .0005$ was observed between CRT FT4 and CRT FT2.

Correlations within VT and CRT were much stronger than correlations between them. This suggests that each has an underlying variable which distinguishes the results of the two assessments, but which is not related to gender or age of the stimulus category. As expected, all the correlations were positive, which fits with the natural expectation that longer VT response times should also correspond to longer CRT response times. Theoretically, it is likely that VT scores are determined by reaction time, while CRT scores are influenced by processing speed.

However, neither seems affected by stimulus age or gender. This is an unexpected finding, and one that will be examined further with the hypothesis testing that follows.

Test-Retest Reliability

Hypothesis 1 was that VT and CRT assessments will be reliable over time. In order to analyse the test-retest reliability, scatterplots were generated for each collapsed category pair for Test 1 and Test 2. These plots are included in Appendix C. One case had significantly longer VT response times in both tests for all categories, also after being Winsorised. Due to this consistency the case was not dropped from other analyses since the data seemed real, but it was removed from the test-retest analysis because it might otherwise contribute to a false positive linear relationship between Test 1 and Test 2. Therefore the sample size for VT was reduced to 50 for the test-retest analysis only. From the distribution of values on the scatterplots it was clear that neither VT nor CRT data showed any signs of correlation between Test 1 and Test 2.

The relationship between Test 1 and Test 2 for both VT and CRT was further investigated using Pearson product-moment correlation coefficient. As expected after the previous examination of the scatterplots, no significant correlations were found between any of the paired categories.

Wilcoxon Signed Ranks tests were done for both the VT and CRT data in order to further investigate the test-retest reliability and compare the response times between the first and second test administrations. Correlation coefficients and test results are included in Appendix D.

The Wilcoxon Signed Ranks test revealed that for the VT data there was a statistically significant difference in the response times for all categories between Test 1 and Test 2 and it was noted that the average response times for each category decreased from Test 1 to Test 2. The largest difference was recorded for the Neutral category from Test 1 (Md = 5.675) to Test 2 (Md = 3.785), z = -4.108, $p < .0005$, with a medium effect size (r = 0.41). The smallest difference was recorded for the MT1 category from Test 1 (Md = 4.960) to Test 2 (Md = 3.985), z = -3.272, $p < .001$, with a medium effect size (r = .33).

For CRT there was no statistically significant difference in the response times for any categories. Again the average response times for each category decreased from Test 1 to Test 2, but only slightly. The largest difference was recorded for the Neutral category from Test 1 (Md = 1.625) to Test 2 (Md = 1.340), z = -1.969, $p < .049$, with a small effect size (r = 0.19). The smallest difference was recorded for the FT3 category from

Test 1 (*Md* = 1.310) to Test 2 (*Md* = 1.175), $z = -1.311$, $p < .190$, with a small effect size ($r = 0.13$).

Interestingly, for CRT there was neither any statistically significant correlation nor any statistically significant difference between Test 1 and Test 2. It was apparent, from the scatter plots in Appendix C, that the CRT values are more tightly concentrated than the VT values but that there was still no apparent relationship between the CRT values for Test 1 and Test 2.

When considered all together, these results indicate that it cannot be concluded that VT and CRT are stable over time, at least not in this study.

Analysis of Variance

The earlier power analysis determined that the resulting group sizes would be too small for a meaningful ANOVA. Therefore a series of Kruskal-Wallis tests were performed to analyse the relationships between the different groups of offenders. Analyses were done on the data from Test 1 to minimise the influence from any possible carryover effects that could otherwise be present in the data from Test 2.

Hypothesis 2a was that sexual deviance should influence VT and CRT results.

Specifically, men who score higher on sexual deviance should have higher VT and CRT scores on more deviant images involving younger children.

To test Hypothesis 2a the participants were grouped according to their sexual deviance as measured with the STABLE-2007. The sexual deviance item is measured as one of either 0 (No indication of sexual deviance), 1 (Some indication of sexual deviance) or 2 (Evidence of sexual deviance). Since only two participants were identified as belonging to the 0 category, all participants with low or no evidence of sexual deviance (scores of 0 and 1) were combined into one group (n=19) , and the high sexual deviance (score of 2) formed the second group (n=32).

Furthermore the VT and CRT categories were grouped into pre-pubescent and post pubescent according to the Tanner categories. This was done by collapsing Tanner categories 1 and 2 (pre-pubescent), excluding Tanner category 3 from the analysis (pubescent) and collapsing Tanner categories 4 and 5 (post pubescent). This was done

separately for female and male categories so the hypothesis could be tested against the resulting four categories: Pre- and post-pubescent female (mean of FT1, FT2 and mean of FT4, FT5 respectively) and pre- and post-pubescent male (mean of MT1, MT2 and mean of MT4, MT5 respectively). Descriptive statistics for the pre-/post-pubescent categories are listed in Appendix E for both VT and CRT.

A Kruskall-Wallis test did not reveal any statistically significant difference in VT scores across the two sexual deviance groups (Gp1, n = 19: Low evidence of sexual deviance, Gp2, n = 32: High evidence of sexual deviance) for any of the categories. Chi-squares (1, n=51) ranged from .277, p = .599 for post-pubescent males (*M* = 5.800, *SD* = 2.008) to 1.190, p = .275 for pre-pubescent males (*M* = 5.461, *SD* = 2.003). Similarly, a Kruskall-Wallis test did not reveal any statistically significant difference in CRT scores across the resulting two sexual deviance groups (Gp1, n = 19: Low evidence of sexual deviance, Gp2, n = 33: High evidence of sexual deviance) for any of the categories. Chi-squares (1, n=52) ranged from .015, p = .902 for pre-pubescent females (*M* = 1.686, *SD* = .709) to .144, p = .704 for post-pubescent females (*M* = 1.759, *SD* = .832). The Kruskall-Wallis statistics are listed in Appendix E, for both VT and CRT.

Hypothesis 2b was that VT and CRT will reliably differentiate child sex offenders who offended only against male victims from those who offended only against female victims.

To test Hypothesis 2b the participants were grouped according to the gender of their victims, resulting in three groups of either female victims, male victims or both. Furthermore the VT and CRT categories were grouped into two categories of female (mean of FT1, FT2, FT3, FT4, FT5) and male (mean of MT1, MT2, MT3, MT4, MT5). Descriptive statistics for the pre-/post-pubescent categories are included in Appendix E, for both VT and CRT.

A Kruskall-Wallis test did not reveal any statistically significant difference in VT scores across the resulting three victim gender groups (Gp1, n = 36: Female victims, Gp2, n = 8: Male victims, Gp3, n = 7: Both) for any of the categories. Chi-squares (2, n=51) was 1.389, p = .499 for females (*M* = 6.092, *SD* = 2.023) and 1.322, p = .499 for males (*M* = 5.689, *SD* = 2.016). Similarly, a Kruskall-Wallis test did not reveal any statistically

significant difference in CRT scores across the resulting three victim gender groups (Gp1, n = 36: Female victims, Gp2, n = 8: Male victims, Gp3, n = 8: Both) for any of the categories. Chi-squares (2, n=52) was .988, p = .610 for female (*M* = 1.690, *SD* = .754) and .328, p = .849 for male (*M* = 1.698, *SD* = .764). The Kruskall-Wallis statistics can be found in Appendix E, for both VT and CRT.

Hypothesis 2c was that VT and CRT will reliably differentiate child sex offenders who offended against prepubescent victims from those who offended against post pubescent victims.

To test Hypothesis 2c the participants were grouped according to the ages of their victims. The participants were divided into three groups composed of those that offended only against victims 12 years old or under, those that offended only against victims 13 years old or over, and those that offended against victims of both age groups. This division also ensured a reasonable number of participants in each group. For testing this hypothesis the same categories were used as for Hypothesis 2a: Pre- and post-pubescent female (mean of FT1, FT2 and mean of FT4, FT5 respectively) and pre- and post-pubescent male (mean of MT1, MT2 and mean of MT4, MT5 respectively). Descriptive statistics for the pre-/post-pubescent categories are included in Appendix E for both VT and CRT.

A Kruskall-Wallis test did not reveal any statistically significant difference in VT scores across the resulting three victim gender groups (Gp1, n = 32: Victims 12 years old or under, Gp2, n = 7: 13 years old or over, Gp3, n = 12: Both) for any of the categories. Chi-squares (2, n=51) ranged from .814, p = .666 for pre-pubescent males (*M* = 5.461, *SD* = 2.003) to 1.515, p = .469 for post-pubescent females (*M* = 6.124, *SD* = 1.978). Similarly, a Kruskall-Wallis test did not reveal any statistically significant difference in CRT scores across the resulting three victim gender groups (Gp1, n = 33: Victims 12 years old or under, Gp2, n = 7: 13 years old or over, Gp3, n = 12: Both) for any of the categories. Chi-squares (2, n=52) ranged from 2.124, p = .346 for post-pubescent males (*M* = 1.697, *SD* = .775) to 2.509, p = .285 for pre-pubescent males (*M* = 1.702, *SD* = .737). The Kruskall-Wallis statistics are listed in Appendix E, for both VT and CRT.

CHAPTER FOUR

DISCUSSION

This study aimed to investigate the utility of the Viewing time and Choice Reaction time assessments as a measure of sexual interest on Child sex offenders in New Zealand. For the purposes of this discussion, each of the hypotheses will be addressed and the findings of the studies discussed with reference to previous research that have used VT and CRT assessments with a similar sample and for a similar purpose. In addition, this discussion will include anecdotal evidence which has potential implications for further studies. This will be followed by a discussion of the limitations and recommendations for future research.

Summary of the Findings

This study found that there was no distinction between the clothed and unclothed stimuli, even though this distinction was supported in the literature (Gress, 2005; Mokros et al., 2010; Wright & Adams, 1994, 1999).One possible explanation for this could be that the constructed images may have not been able to elicit the individuals' subjective perception of what they considered to be sexually attractive. The quality of the constructed images and participants' awareness that the images were constructed may have also had an influence in terms of what they perceived to be real or not. In addition, due to the overrepresentation of Maori participants in the study, their sexual interest may have not been adequately represented through the images, since all of them were of Caucasian people.

Hypothesis one; that VT and CRT assessments will be reliable over time was not supported by this study. Participants' responses to Test 1 and Test 2 demonstrated that there was no statistically significant correlation between the tests. In addition, it was found that the average response times decreased from Test 1 to Test 2. As of October 2011 there are no other current studies that assess the test-retest reliability of the VT and CRT assessments and therefore the findings of this study cannot be compared with others. Given that this is an exploratory study and that test-retest has not been done

before, this study provides valuable information on the reliability of the VT and CRT assessments over time.

The lack of support for the test-retest reliability of the VT and CRT assessments could be due to several reasons. The general decrease in response times from Test 1 to Test 2 indicates the presence of carryover and learning effects. Participants may have learned to read and answer the questions faster (e.g. they became faster at locating the dots on the CRT assessment). It is possible that the time between Test 1 and Test 2 may not have been long enough to minimize the effects of the first test on the second test.

The participants may also already have become sensitized to the type of changes they need to make with regards to sexual interest towards children. Therefore they could have become more aware of their responses the second time the assessments were administered. Child sex offenders who attend the Te Piriti STU programme may also have already attended other treatment programmes while in prison, or in the community, and may have started working on the management of their sexual interest towards children. In addition, once in the treatment unit, participants may become more aware of additional strategies on how to manage their sexual interest or arousal towards children by discussing it or hearing it from other prisoners. For Test 1, confounding variables such as the ones mentioned above were managed by assessing participants soon after they had arrived to the unit; typically within a couple of weeks. For Test 2, the assessments took place three to four weeks after Test 1, since participants were at that time still attending the preparatory stages of treatment and therefore had less exposure to information about the core stage of the programme. It is still possible that previous treatment and information about treatment, as well as the therapeutic environment in itself, acted as a confounding variable during the VT and CRT assessments. However, those confounding variables are to be expected within the population of offenders attending treatment and it is likely that any assessment performed on such will have to face similar circumstances.

Finally, the length of the VT (415 items) and CRT (425 items) assessments may have had an influence on participants' attitudes and motivation to respond to the assessments, especially when they were given the entire test a second time. Both the VT and CRT assessments were administered consecutively on the same day. Anecdotal evidence showed that participants commented that the test was too long and that they were bored

having to look at the same pictures repeatedly. A couple of participants commented on the speed in which they answered the VT questions, since they wanted to complete the assessment rapidly. Participants' comments can be found in Appendix F. Gress (2007) highlighted similar issues with the length of the assessment. The rationale behind the many items of the assessment is that by pairing every image with five questions each, it would balance the influence of the question on the image, such as how long it would take the participant to answer the question (Gress, 2007).

Hypothesis 2a was that sexual deviance should influence VT and CRT results. Specifically, men who score higher on sexual deviance should have higher VT and CRT scores on more deviant images involving younger children. There are no current studies that have compared VT and CRT assessments with a risk assessment measure such as the STABLE-2007 and therefore this study's findings cannot be compared with other studies.

Results indicated that no statistically significant differences were found for participants' VT and CRT scores when grouped according to the evidence of sexual deviance as determined by the STABLE-2007. Even though there was no support for this hypothesis, some interesting tendencies were noticed when analysing the data. Participants from the group with evidence of sexual deviance (HSD) showed shorter viewing times than participants from the group with no or low evidence of sexual deviance (LSD). This was reversed in the CRT assessment where participants from the HSD group took longer time to answer the questions than participants from the LSD group. These findings could indicate that VT questions had an important effect on participant's responses. VT assessment gives the impression that it is a self-report assessment and the literature indicates that offenders are less willing to disclose their deviant sexual preferences due to potential legal repercussions (Hunter, Becker, & Kaplan, 1995; Laws, Hanson, Osborn, & Geenbaum, 2000; Seto, 2008; Seto, 2009). Anecdotally, some participants commented on the conscious avoidance of having to look at the images and therefore answered questions rapidly. It can be assumed that participants may have been more aware of their responses and focused mainly on providing a fast answer to the questions rather than spending time looking at the images.

CRT assessments on the other hand are different, since the alternative task of having to

locate a dot distracts from the concerns of revealing any sexual interest to children. The instructions for the CRT assessment also state that participants should answer the questions as fast as possible. Participants in the HSD group may thus have become less aware of their sexual preferences during the CRT assessment and looked at the pictures longer, since they were less concerned about what or how to respond. In this context it is noteworthy to consider that Te Piriti STU is a treatment unit and every offender who decides to attend the programme is fully aware that their participation in treatment overall will have an impact on any report written about them. This may have an influence on their level of compliance and attitude towards assessment and activities related to treatment. On the other hand, some offenders may be more willing to undertake assessments since they would like to understand their offending and place safety measures around them in order to avoid reoffending.

Hypothesis 2b was that VT and CRT will reliably differentiate child sex offenders who offended only against male victims from those who offended only against female victims.

In contrast to the Gress (2005) study that found that VT could differentiate child sex offenders' gender preference, the results found in this study did not indicate any statistically significant differences in the VT and CRT scores in relation to the gender of the victims.

A closer analysis of the data revealed that there seems to be a tendency for participants who offended against both male and female victims to have longer response times in both the VT and CRT assessments compared to those who only offended against either males or females. The literature (Hanson & Bussiere, 1998) shows that having more than one victim and having male victims is related to the presence of sexual deviance. It is possible that the findings for this study indirectly indicate such a presence. The present study had however a small sample and the statistical significance was not high enough for any definite conclusions to be made.

Hypothesis 2c was that VT and CRT will reliably differentiate child sex offenders who offended against prepubescent victims from those who offended against post pubescent victims.

The data analysis did not reveal statistically significant differences in the VT and CRT scores in relation to the age of victims. Both Gress (2005) and Mokros et al., (2010)

found that VT and CRT assessments respectively could differentiate child sex offenders who offended against prepubescent victims and that their viewing times were longer than when looking at images of adults.
Despite the lack of statistical significance for this hypothesis, participants for this study who had offended only against children of age 12 years and younger had longer response times both in the VT and CRT assessments.

An important aspect to consider is that this study included images of people who were both clothed and nude. Even though the literature supports the use of both types of stimuli (Gress, 2005; Mokros et al., 2010; Wright & Adams, 1994, 1999), there is no current literature on the influence of culture on these types of assessments. Maori offenders were overrepresented in this study. Maori participants' comments revealed that many found images of nude children particularly uncomfortable. In addition, some of them expressed their concern in terms of their cultural values. One participant stated that looking at images of nude children was Tapu (forbidden). In addition, another participant expressed the need to do a Karakia (prayer) after having completed the assessment. These issues may have had an impact on the motivation, attitude and performance of these participants in both the VT and CRT assessments. Presentation of these images to Maori offenders were however discussed with a Maori Cultural Consultant who found that the assessments could provide relevant information into the assessment of sexual offenders and therefore the VT and CRT administration was justified.

It should be noted that the Cultural Consultant advised that it was possible that offenders in such cases might be referring to their culture as an excuse for not addressing their sexual offending against children. For example, the Cultural Consultant advised that the use of a term such as Tapu within the context of having sexually offended against a child showed that the offenders had a lack of understanding of the meaning and proper use of the term. Instead a Maori person who is fully aware of the Maori protocols would make proper use of Karakia, meaning that the appropriate process would be to do a Karakia at the beginning and at the end (M. Webb, personal communication, October 11, 2011).

Furthermore, some participants identified themselves as computer illiterate and were afraid of making mistakes. This could also have impacted their response times.

Limitations

The current study had a number of noteworthy limitations and the findings need to be considered in light of these.

The sample size for the study was small. A larger sample size would have strengthened the results and would have enabled a more powerful statistical analysis. However, only 52 child sex offenders were considered suitable to be assessed with the VT and CRT in the period allocated for testing at Te Piriti Special Treatment Unit.

The sample for this study did not adequately represent the ethnic distribution of general child sex offenders who attend treatment. European child sex offenders represent the majority of those who attend treatment; however in this study Maori offenders (53.8%) were overrepresented. This could be particularly important because the constructed images contained in the VT and CRT assessments are only of Caucasian people. This may have been an issue when assessing Maori participants with this particular VT and CRT test, since people on average may be more inclined to be sexually attracted to people of their own ethnicity. Anecdotally, a couple of Maori participants commented that if the assessments would have included people of darker skin colour, their sexual interest towards the images would have been greater. Furthermore, only males were available to be assessed with the VT and CRT assessments in this study and therefore this limits the generalisation of the results to males only.

The age distribution of the general offender victims was not adequately represented in this study. Participants for this study had a higher percentage of prepubescent victims only (younger than 12 years old). It is possible that participants may have been aware of the negative implications of their sexual attraction to prepubescent children in terms of their risk of re-offending. This may have acted as a confounding variable, since participants may have been inclined to show less sexual interest towards this age group when being assessed. The impact of this is believed to be minimal however, since the assessments record viewing times in milliseconds and therefore is less prone to be faked.

Due to the prison setting, personal and emotional circumstances may have had an

influence on participants' attitudes and responses to the assessments. The assessor could not control for incidents or changes that occurred in the prison environment on the day. Participants may have been exposed to arguments with other prisoners, disagreements with officers or may have received unpleasant personal news from their families. Personal and emotional variables such as the above could not be controlled for. However, each participant was given the opportunity to discuss general issues associated with the assessment prior to being assessed. The VT and CRT assessments were administered on participants on different days and different times of the day, depending on the assessor's availability. This also applied for both Test 1 and Test 2. Participants may have been less willing to respond to the assessment on the day and time they were called by the assessor.

The research setting might also have been more relaxed if the room used would have not been the same as the one used for the Plethysmograph assessment. Child sex offenders attending the Te Piriti STU treatment programme are often predisposed to view the Plethysmograph assessment as negative. Participants may have been expecting to be assessed with the Plethysmograph assessment and their anxiety levels may have had an influence on their performance on the assessments. On the other hand, participants may have been more willing to do the VT and CRT assessments since it appeared less intrusive and they might have felt a greater sense of control when responding to the questions.

Future Directions

A replication of the current study would be valuable, given the lack of research on VT and CRT assessments in New Zealand. Any replication should include a larger sample size to allow for a more powerful statistical analysis. It is suggested that the collection of data should be done over a longer period of time than one year, since it will allow for a better random selection of the sample.

This study did not include a control group, which would have added to the strengths of the findings. Participants for this study acted as their own control group, since all participants were administered Test 1 and Test 2 at two points in time. It would be ideal to have a group of non sex offenders to compare the participants against in terms of

their responses and the time spent viewing at the images for the purpose of adding to the discriminative validity of the assessments. However, due to ethical concerns with regards to the emotional impact the VT and CRT assessments could have had on non-sexual offenders, the access of a control group was not possible.

It would also be useful to use a Social Desirability scale. Several studies have shown that offenders may be less willing to disclose their deviant sexual preferences due to potential legal repercussions (Hunter, Becker, & Kaplan, 1995; Laws, Hanson, Osborn, & Geenbaum, 2000; Seto, 2008; Seto, 2009). It was assumed that social desirable responses to the assessments were minimised due to participants being unaware that the response times were being recorded, as well as due to the millisecond accuracy. This can potentially still have influenced the response times somewhat though.

VT and CRT assessments should be administered independently if they are to be used in its original long form. Participants may then be less prone to boredom and their attitude towards the test may improve.

With regards to the test-retest of the VT and CRT assessments, even though there are no specific timeframes with regards to when Test 2 should be administered (Hogan, 2003; Kaplan & Saccuzzo, 2005; Kline, 2005), it may be appropriate to ensure that participants have minimal exposure to treatment information in between tests. In addition, in order to minimize carryover and learning effects from the first test onto the second, VT and CRT assessments could be presented in a different order (e.g. VT first on test 1 and CRT first on test 2). Furthermore, it will be worthwhile to explore whether the test-retest reliability of the assessments improve if the items within the assessments are presented in a different or random order the second time.

Limestone Technologies Inc. (2011) has recently developed new software to assess Visual Sexual Preference. This appears to be an improved version in comparison to the software that was used for this study. Among some of the improvements, participants are asked to estimate the age of the person they are viewing, and this age will be shown on top of the page the next time they look at the image. Participants' ability to estimate the age of a person provides important clinical information with regards to their sexual preference. In addition, the new software contains improved images and includes both Caucasian and African American models, which makes the assessment of people with

different ethnicities more meaningful.

It is recommended that the Department of Corrections continue research into the utility of these types of assessments. Even though the combined VT/CRT assessment was long it was still shorter than the typical 2 hours required for a Plethysmograph assessment. When considering that the VT/CRT is also less intrusive and do not require specialised training or equipment it would be highly beneficial for the Department of Corrections to continue the research into the utility of these measures.

As part of the standard programme in Te Piriti the new Visual Sexual Preference assessment from Limestone Technologies could be used on a regular basis so that a sufficiently large sample is collected over time in order to facilitate further analysis. This new assessment is also shorter than the VT/CRT used for this study and therefore the issues with boredom can be minimised. Also, with an overall shorter assessment time the collection of a larger sample is more realistic. The improved images also address the limitations regarding different ethnicities to a degree so the assessment might be better suited for the offender population at Te Piriti.

Conclusion

This study did not find support for the utility of Viewing time and Choice reaction time assessments with child sex offenders in New Zealand. The study did however uncover a number of important limitations of the VT/CRT software itself and the method of assessment, which have not been addressed previously in literature. Research into the use of these assessments, especially with a New Zealand population of child sex offenders, is in its infancy and therefore further studies are needed. A list of suggestions has been given for further research and it is important that the limitations outlined here are addressed. This is particularly important because the findings of this study could have a significant influence on any decision to use VT/CRT assessments within the Department of Corrections.

REFERENCES

Abel Screening Incorporated (2011). *Tests from Abel Screening*. Retrieved 7 August, 2011, from http://www.abelscreening.com/tests/index.html

Abel, G. G. (1985). *A clinical evaluation of possible sex offenders. The incest offender, the victim, the family*. New York: Mental Health Association of Westchester County Inc.

Abel, G. G. (2000). The Importance of Meeting Research Standards: a Reply to Fischer and Smith's Articles On the Abel Assessment for Sexual Interest TM. *Sexual Abuse: A Journal of Research and Treatment, 12*, 155-161.

Abel, G. G., Becker, J. V., Mittelman, M. S., Cunningham-Rathner, J., Rouleau, J. L., & Murphy, W. D. (1987). Self-reported sex crimes of nonincarcerated paraphiliacs. *Journal of Interpersonal Violence, 2,* 3 - 25.

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*(2), 81-95.

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

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,* 255-265.

Aboyoun, D. C., & Dabbs, J. M. J. (1998). The Hess pupil dilation findings: Sex or novelty?. *Social Behavior and Personality, 26*(4), 415-419.

Abramson, P. R., Perry, L. B., Seeley, T. T., Seeley, D. M., & Rothblatt, A. B. (1981). Thermographic measurement of sexual arousal: A discriminant validity analysis. *Archives of Sexual Behavior, 10,* 171–176.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multimethod approach. *Criminal Justice and Behavior, 37*(3), 319-335.

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. C*anadian Journal of Behavioural Science, 21*(1), 70-82.

Beck, J. G., Barlow, D. H., & Sakheim, D. K. (1983). Abdominal Temperature Changes During Male Sexual Arousal. *Psychophisiology, 20*(6), 715-717.

Beech, A. R. (2001). Case material and interview. In C. Hollin (Ed.), *Handbook of offender assessment and treatment* (pp. 123-138). Chichester: Wiley.

Blanchard, R. (2009). Reply to Letters Regarding *Pedophilia, Hebephilia and the DSM-V. Archives of Sexual Behavior, 38,* 331-334.

Blanchard, R., Lykins, A. D., Wherrett, D., Kuban, M. E., Cantor, J. M., Blak, T., …Klassen, P. E. (2009). Pedophilia, Hebephilia, and the *DSM-V. Archives of Sexual Behavior, 38,* 335-350.

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

Card, R. D., & Farrall, W. R. (1990). Detecting faked penile responses to erotic stimuli: A comparison of stimulus conditions and response measures. *Annals of Sex Research, 3*, 381–396.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Connolly, M. (2004). Developmental Trajectories and Sexual Offending: An Analysis of the Pathways Model. *Qualitative Social Work: Research and Practice, 3*(1), 39-59.

Côté, K., Earls, C. M., & Lalumière, M. L. (2002). Birth Order, Birth Interval, and Deviant Sexual Preferences Among Sex Offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*(1), 67-81.

Cox, J. (2006). Review of Profiling violent crimes: An investigative tool. *Journal of Investigative Psychology and Offender Profiling, 3*(2), 134-137.

Craig, L. A., Browne, K. D., Beech, A., & Stringer, I. (2006). Psychosexual characteristics of sexual offenders and the relationship to sexual reconviction. *Psychology, Crime & Law, 12*(3), 231-243.

Dandescu, A., & Wolfe, R. (2003). Considerations on Fantasy Use by Child Molesters and Exhibitionists. *Sexual Abuse: A Journal of Research and Treatment, 15*(4), 297-305.

Dowling, N., Smith, D., Proeve, M., & Lee, J. K. P. (2000). The Multiphasic Sex Inventory: A Comparison of American and Australian Samples of Sexual Offenders. *Australian Psychologist, 35*(3), 244-248.

Earls, C. M. (1983). Some Issues in the Assessment of Sexual Deviance. *International Journal of Law and Psychiatry, 6*, 431-441.

Elger, B. S. (2008). Research Involving Prisoners: Consensus and Controversies in International and European Regulations. *Bioethics, 22*(4), 224–238.

Eyler, L. T., & Jeste, D. V. (2006). Enhancing the Informed Consent Process: A Conceptual Overview. *Behavioral Sciences and the Law, 24*, 553-568.

Fanslow, J., Robinson, E. M., Crengle, S., & Perese, L. (2007). Prevalence of child sexual abuse reported by a cross-sectional sample of New Zealand women. *Child Abuse & Neglect, 31*(9), 935-945.

Farrall, W. R., & Card, R. D. (1988). Advancements in physiological evaluation of assessment and treatment of the sexual aggressor. *Annals of the New York Academy of Sciences, 528*, 261–273.

Fergus, L., & Keel, M. (2005). *Adult victim/survivors of childhood sexual assault.* Retrieved September 9, 2011, from www.aifs.gov.au/acssa/pubs/wrap/acssa_wrap1.pdf

Finkelhor, D. (1984). *Child Sexual Abuse: New Theory & Research.* New York: Free Press.

Fischer, L., & Smith, G. (1999). Statistical Adequacy of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11,* 195-205.

Flak, V., Beech, A., Fisher, D. (2007). Forensic assessment of deviant sexual interests: The current position. *Issues in Forensic Psychology, 6,* 70-83.

Fleming, T. M., Watson, P. D., Robinson, E., Ameratunga, S., Dixon, R., Clark, T.C., & Crengle, S. (2007). *Violence and New Zealand Young People: Findings of Youth2000 – A national secondary school youth health and wellbeing survey*. The University of Auckland, Auckland.

Freund, K. (1963). A Laboratory Method For Diagnosing Predominance Of Homo- Or Hetero- Erotic Interest In The Male. *Behaviour Research and Therapy, 1,* 85-93.

Freund, K. (1965). Diagnosing Heterosexual Pedophilia By Means Of A Test For Sexual Interest. *Behaviour Research and Therapy, 3,* 229-234.

Freund, K. (1967). Diagnosing Homo- Or Heterosexuality And Erotic Age-Preference By Means Of A Psychophysiological Test. *Behaviour Research and Therapy, 5,* 209-228.

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: A Journal of Consulting and Clinical Psychology, 3*(2), 254-260.

Freund, K., & Watson, R. J. (1992). The Proportions of Heterosexual and Homosexual Pedophiles Among Sex Offenders Against Children: An Exploratory Study. *Journal of Sex & Marital Therapy, 18*(1), 34-43.

Fuller, A. K., Barnard, G., Robbins, L., Spears, H. (1988). Sexual maturity as a criterion for classification of phallometric stimulus slides. *Archives of Sexual Behavior, 17*(3), 271-276.

Garrett, J. C., Harrison, D. W., & Kelly, P. L. (1989). Pupillometric Assessment of Arousal to Sexual Stimuli: Novelty Effects or Preference?. *Archives of Sexual Behavior, 18*(3), 191-201.

Glasgow, D. V. (2009). Affinity: The Development Of A Self-Report Assessment Of Paedophile Sexual Interest Incorporating A Viewing Time Validity Measure. In D. Thornton & D. R. Laws (Eds.), *Cognitive Approaches To The Assessment Of Sexual Interest IN Sexual Offenders* (pp. 59-84). Wiley-Blackwell

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

Goldwater, B. C. (1972). Psychological Significance Of Pupillary Movements. *Psychological Bulletin*, *77*(5), 340-355.

Gray, N. S., Brown, A. S., MacCulloch, M. J., Smith, J., & Snowden, R. J. (2005). An Implicit Test of the Associations Between Children and Sex in Pedophiles. *Journal of Abnormal Psychology, 114* (2), 304–308.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to Measure

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464-1480.

Gress, C. L (2007). *Delays in attentional processing when viewing sexual imagery: The development and comparison of two measures* (Doctoral dissertation, University of Victoria). Retrieved from http://dspace.library.uvic.ca:8080/bitstream/handle/1828/1234/Gress%20Dissertation%20FINAL%20May%2014-07.pdf

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

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

Groth, A. N. & Birnbaum, H. J. (1978). Adult sexual orientation and attraction to underage persons. *Archives of Sexual Behavior, 7*(3), 175–181.

Grubin, D. (2008). Medical Models And Interventions In Sexual Deviance. In R. D. Laws & W. T. O'Donohue (Eds.), *Sexual Deviance: Theory, assessment and treatment* (pp. 594-610). New York, NY: Guilford Press.

Hall, G. C. N., & Hirschman, R. (1992). Sexual aggression against children: A conceptual perspective of etiology. *Criminal Justice and Behavior, 19,* 8-23.

Hamel, R. F. (1974). Female subjective and pupillary reaction to nude male and female figures. *The Journal of Psychology, 87,* 171-175.

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or malleability of the IAT?. *Journal of Experimental Social Psychology*, *46*, 286-298.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally-created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology*, *42*, 259-272.

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

Hanson, R. K., & Harris, A. (1998). *Dynamic Predictors Of Sexual Recidivism.* (User Report 1998-01). Ottawa: Department of the Solicitor General of Canada.

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene?. Dynamic predictors of sex offense recidivism. *Criminal Justice and Behavior, 27,* 6–35.

Hanson, R. K., & Morton-Bourgon, K. (2004). Predictors of Sexual Recidivism: An Updated Meta-Analysis. Public Works and Government Services Canada.

Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (User Report No 2007–05). Ottawa, Ontario: Public Safety Canada.

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.* User Report, Corrections Research, Ottawa: Public Safety Canada.

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*(3), 296-310.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing Time as a Measure of Sexual Interest Among Child Molesters And Normal Heterosexual Men. *Behaviour Research and Therapy, 34*(4), 389-394.

Hess, E. H., & Polt, J. M. (1960). Pupil Size as Related to The Interest Value of Visual Stimuli. *Science, 132,* 349–350.

Hinton, J. W., O'Neill, M. T., & Webster, S. (1980). Psychophysiological assessment of sex offenders in a security hospital. *Archives of Sexual Behavior, 9,* 205–216.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmidt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin 31*(10), 1369-1385

Hogan, T. P. (2003). Reliability. In T. P. Hogan (Ed.), *Psychological Testing: A Practical Introduction* (pp. 123-170). New York: John Wiley & Sons.

Holland, L. A., Zolondek, S. C., Abel, G. G., Jordan, A. D., & Becker, J. V. (2000). Psychometric analysis of the Sexual Interest Cardsort Questionnaire. *Sexual Abuse: Journal of Research and Treatment, 12,* 107–122.

Holmes, R. M., & Holmes, S. T. (2009). *Profiling violent crimes: An investigative tool.* Thousand Oaks, CA: Sage Publications.

Howard, R. C., Longmore, E. J., Mason, P. A., & Martin, J. L. (1994). Contingent negative variation (CNV) and erotic preference in self-declared homosexuals and in child sex offenders. *Biological Psychology, 38,* 169-181.

Hunter, J. A., Becker, J. V., & Kaplan, M. S. (1995). The Adolescent Sexual Interest Card Sort: Test-Retest Reliability and Concurrent Validity in Relation to Phallometric Assessment. *Archives of Sexual Behavior, 24*(5), 555-561.

Imhoff, R., Schmidt, A. F., Nordsiek, 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.

InnocentDads.org. (2002). *Explaining the Abel Assessment*. Retrieved 15 November, 2008, from http://www.innocentdads.org/abel.htm

Israel, E., & Strassberg, D. S. (2009). Viewing Time as an Objective Measure of Sexual Interest in Heterosexual Men and Women. *Archives of Sexual Behavior, 38*(4), 551-558.

Janisse, M. P. (1977). Pupillometry: The Psychology of the Pupillary Response. Washington, DC: Hemisphere Publishing Co.

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*(4), 384-396.

Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*, 193–217.

Kaplan, R. M., & Saccuzzo, D. P. (2005). Reliability. In R. M. Kaplan & D. P. Saccuzzo (Eds.), *Psychological Testing: Principles, Applications, and Issues* (pp. 99-131). Belmont, CA: Thomson Wadsworth.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*(5), 774-788.

Kline, T. J. B. (2005). Reliability of Test Scores and Test Items. In T. J. B. (Ed.), *Psychological Testing: A Practical Approach to Design and Evaluation* (pp. 167-184). Thousand Oaks, California: Sage Publications.

Krueger, R. B., Bradford, J. M. W., & Glancy, G. D. (1998). Report from the Committee on Sex Offenders: The Abel Assessment for Sexual Interest-A Brief Description. *Journal of the American Academy of Psychiatry and the Law, 26,* 277-278.

Lalumiere, M. L., Harris, G. T., Quinsey, V. L., Rice, M. E. (1998). Sexual Deviance and Number of Older Brothers Among Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*(1), 5-15.

Larson, J., Robertson, P., Hillman, D., & Hudson, S. (1998). Te Piriti: A bicultural model for treating child molesters in Aotearoa / New Zealand. In W. Marshall, S. M. Hudson, T. Ward, & Y. Fernandez (Eds.), *Sourcebook of treatment programs for sexual offenders* (pp. 385-398). New York: Plenum.

Laws, D. R., & Gress, C. L. Z. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9,* 1-14.

Laws, D. R., & Marshall, W. L. (1991). Masturbatory reconditioning with sexual deviates: an evaluative review. *Advanced Behaviour Research & Therapy, 13,* 13-25.

Laws, D. R., Hanson, R. K., Osborn, C. A., & Geenbaum, P. E. (2000). Classification of Child Molesters by Plethysmographic Assessment of Sexual Arousal and a Self-Report Measure of Sexual Preference. *Journal of Interpersonal Violence, 15*(12), 1297-1312.

Laws, D.R. (1995). Verbal satiation: notes on procedure with speculations on its mechanism of effect. *Sexual Abuse: A Journal of Research and Treatment, 7,* 155–166.

Leitenberg, H., & Henning, K. (1995). Sexual Fantasy. *Psychological Bulletin, 117*(3), 469-496.

Letourneau, E. J. (2002). A Comparison of Objective Measures of Sexual Arousal and Interest: Visual Reaction Time and Penile Plethysmography. *Sexual Abuse: A Journal of Research and Treatment, 14*(3), 207-223.

Levay, S., & Valente, S. (2006). Sexual Attraction and Arousal. In S. Levay, & S. Valente (Eds.), *Human Sexuality* (pp. 216-251). Sunderland, MA: Sinauer Associates, Inc.

Limestone Technologies Inc. (2011). Visual Sexual Preference Assessment. Retrieved 11 October, 2011, from http://www.limestonetech.com/index.cfm/product-lines/sex-offender-management/visual-sexual-preference-assessment/

Looman, J., Abracen, J., Maillet, G., & DiFazio, R. (1998). Phallometric Nonresponding in Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*(4), 325-336.

Love, R. E., Sloan, L. R., & Schmidt, M. J. (1976). Viewing Pornography and Sex Guilt: The Priggish, the Prudent, and the Profligate. *Journal of Consulting and Clinical Psychology, 44*(4), 624-629.

Lussier, P., Beauregard, E., Proulx, J., & Nicole, A. (2005). Developmental Factors Related to Deviant Sexual Preferences in Child Molesters. *Journal of Interpersonal Violence, 20*(9), 999-1017.

Marquis, J. N. (1970). Orgasmic reconditioning: Changing sexual object choice through controlling masturbation fantasies. *Journal of Behavior Therapy and Experimental Psychiatry, 1*, 263-271.

Marshall, W. L. & Fernandez Y. M. (2003). Sexual preferences: Are they useful in the assessment and treatment of sexual offenders?. *Aggression and Violent Behavior, 8,* 131-143.

Marshall, W. L. (2006). Clinical and research limitations in the use of phallometric testing with sexual offenders. *Sexual Offender Treatment, 1*(1), 1-18.

Marshall, W. L., & Barbaree, H. E. (1990). An Integrated Theory of the Etiology of Sexual Offending. In W. L. Marshall, D. R. Laws, & H. E. Barbaree (Eds), *Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 257-275). New York: Plenum.

Marshall, W. L., & Eccles, A. (1991). Issues in Clinical Practice With Sex Offenders. *Journal of Interpersonal Violence, 6*(1), 68-93.

Marshall, W. L., (1997). The relationship between self esteem and deviant sexual arousal in nonfamilial child molester. *Behavior Modification, 21,* 86-96.

Marshall, W. L., Barbaree, H. E., & Eccles, A. (1991). Early Onset and Deviant Sexuality in Child Molesters. *Journal of Interpersonal Violence, 6*(3), 323-336.

Marshall, W.L. (1979). Satiation therapy: a procedure for reducing deviant sexual arousal. *Journal of Applied Behavioral Analysis, 12,* 10–22.

McGuire, R. J., Carlisle, J. M., & Young, B. G. (1965). Sexual Deviations as Conditioned Behaviour: A Hypothesis. *Behaviour Research and Therapy, 2*(2), 185-190.

Merdian, H. L., & Jones, D. T. (2011). Phallometric Assessment of Sexual Arousal. In D. P. Boer, R. Eher, L. A. Craig, M. H. Miner, F. Pfäfflin (Eds.), *International Perspectives on the Assessment and Treatment of Sexual Offenders: Theory, Practice and Research* (pp. 141-169). Chichester, West Sussex: Wiley-Blackwell.

Mihailides, S., Devilly, G. J, Ward, T. (2004). Implicit cognitive distortions and sexual offending. *Sexual Abuse: A Journal of Research and Treatment, 16*(4), 333-350.

Miller, J. (1991). Reaction Time Analysis with Outlier Exclusion: Bias Varies with Sample Size. The Quarterly Journal of Experimental Psychology, 43A(4), 907-912.

Mokros, A., Dombert, B., Osterheider, M., Zappala, A., & Santtila, P. (2010). Assessment of Pedophilic Sexual Interest with an Attentional Choice Reaction Time Task. *Archives of Sexual Behavior, 39,* 1081-1090.

Mullen, P. E., King, N. J., Tonge, B. J. (2000). Child Sexual Abuse: An Overview. *Behaviour Change, 17,* 2-14.

Mullen, R., Anderson, J., Roman-Clarkson, S., & Martin, J. (1991). *Otago Women's Health Survey.* Unpublished manuscript, Otago University Medical School, Dunedin.

Murphy, W. D., & Barbaree, H. E. (1994). *Assessments of sex offenders by measures of erectile response: Psychometric properties and decision making.* Victoria: Safer Society Press.

Nunes, K. L., Firestone, P., & Baldwin, M. W. (2007). Indirect Assessment of Cognitions of Child Sexual Abusers with the Implicit Association Test. *Criminal Justice and Behavior, 34*(4), 454-475.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86,* 653-667.

Parole Act 2002 No10, §107a (2004).

Parole Act 2002 No10, §15a, §33.

Price, S. A., & Hanson, K. R. (2007). A modified Stroop task with sexual offenders: Replication of a study. *Journal of Sexual Aggression, 13*(3), 203-216.

PsyPhy (1999). *Age and Gender Exemplars* {CD-ROM}. Author: Halifax, NS, USA.

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

Quinsey, V. L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing Time as a Measure of Sexual Interest. *Ethology and Sociobiology, 17,* 341-354.

Rathus, S. A., Nevid, J. S., & Fichner-Rathus, L. (2008). Attraction and Love-Binding Forces. In S. A. Rathus, J. S. Nevid, & L. Fichner-Rathus (Eds.), *Human Sexuality: In a World of Diversity* (pp. 196-221). New York: Pearson.

Rosenzweig, S. (1942). The Photoscope As An Objective Device For Evaluating Sexual Interest. *Psychosomatic Medicine, 4,* 150-158.

Rothermund, K., Wentura, D. (2004). Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations. *Journal of Experimental Psychology: General, 133*(2), 139–165.

Rupp, H. A., & Wallen, K. (2008). Sex Differences in Response to Visual Sexual Stimuli: A Review. *Archives of Sexual Behavior, 37,* 206-218.

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

Scott, T. R., Wells, W. H., & Wood, D. Z. (1967). Pupillary Response and Sexual Interest Reexamined. *Journal of Clinical Psychology,.23*(4), 433-438.

Self-Esteem and Self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022-1038.

Sentencing Act 2002 No 9, §87.

Seto, M. C. (2001). The Value of Phallometry in the Assessment of Male Sex Offenders. *Journal of Forensic Psychology Practice, 1*(2), 65-75.

Seto, M. C. (2008). Assessment Methods. In M. C. Seto (Ed.), *Pedophilia and sexual offending against children: Theory, assessment and intervention* (pp. 23-45). Washington, DC: American Psychological Association.

Seto, M. C. (2008). Pedophilia: Psychopathology and Theory. In R. D. Laws & W. O'Donohue (Eds.), *Sexual Deviance: Theory, assessment and treatment* (pp. 164-182). New York: Guilford Press.

Seto, M. C. (2009). Pedophilia. *Annual Review Clinical Psychology, 5,* 391-407.

Simkins, L., Ward, W., Bowman, S., & Rinck, C. M. (1989). The Multiphasic Sex Inventory: Diagnosis and prediction of treatment response in child sexual abusers. *Annals of Sex Research, 2*, 205–226.

Singer, B. (1984). Conceptualizing Sexual Arousal and Attraction. *The Journal of Sex Research, 20,* 230-240.

Smith, G., & Fischer, L. (1999). Assessment of Juvenile Sexual Offenders: Reliability and Validity of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: A Journal of Research and Treatment, 11,* 207-216.

Smith, P., & Waterman, M. G. (2004) Processing Bias for Sexual Material: the Emotional Stroop and Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*(2), 163-171.

Snowden, R. J., Craig, R. L., & Gray, N. S. (2011). Indirect Behavioral Measures of Cognition among Sexual Offenders. *Journal of Sex Research, 48*(2-3), 192-217.

Steffens, M. C., Yundina, E., & Panning, M. (2008). Automatic Associations with "Erotic" in Child Sexual Offenders: Identifying Those in Danger of Reoffence. *Sexual Offender Treatment, 3*(2).

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662.

Swaffer, T., Hollin, C., Beech, A., Beckett, R., & Fisher, D. (2000). An Exploration of Child Sexual Abusers' Sexual Fantasies Before and After Treatment. *Sexual Abuse: A Journal of Research and Treatment, 12*(1), 61-68.

Tanner, J. M. (1962). *Growth at Adolescence*. Oxford: Blackwell Scientific.

Van Hecke, T. (2010). Power study of anova versus Kruskal-Wallis test. Retrieved March 10, 2011, from http://interstat.statjournals.net/YEAR/2010/articles/1011002.pdf

Van Leeuwen, M., van Baaren, R., Chakhssi, F., Loonen, M.,Lippman, M., & Dijksterhuis, A. (2009). *Detecting implicit paedophilic preferences: Improving predictability*. Manuscript submitted for publication.

Ward, T. (2002) Marshall and Barbaree's Integrated Theory of Child Sexual Abuse: A Critique. *Psychology, Crime & Law, 8*, 209-228.

Ward, T., & Siegert, R. J. (2002). Toward a comprehensive theory of child sexual abuse: A theory knitting perspective. *Psychology, Crime & Law, 8*(4), 319-351.

Ward, T., & Sorbello, L. (2003). Explaining Child Sexual Abuse: Integration and Elaboration. In T. Ward, D. R. Laws, & S. M. Hudson (Eds.), *Sexual Deviance: Issues and Controversies* (pp. 3-20). Thousand Oaks, California: Sage Publications.

Ward, T., Hudson, S. M., Marshall, W. L.,& Siegert, R. (1995). Attachment style and intimacy deficits in sexual offenders: A theoretical framework. *Sexual Abuse: A Journal of Research and Treatment, 7*(4), 317-335.

Webster, J. S., & Hammer, D. (1983). Thermistor measurement of male sexual arousal. *Psychophysiology, 20*, 111–115.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*(3) , 475-482.

Wieckowski, E., Hartsoe, P., Mayer, A., & Shortz, J. (1998). Deviant Sexual Behavior in Children and Young Adolescents: Frequency and Patterns. *Sexual Abuse: A Journal of Research and Treatment, 10*(4), 293-303.

Williams, K. M. (2003). Two techniques for assessment of sexual interest: A discussion of the clinical utility of penile plethysmography and visual reaction time. *The Forensic Examiner*, 35-38.

Worling, J. R. (2006). Assessing Sexual Arousal with Adolescents Males Who Have Offended Sexually: Self Report and Unobtrusively Measured Viewing Time. *Sexual Abuse: A Journal of Research and Treatment, 18*, 383-400.

Wormith, J. S. (1986). Assessing Deviant Sexual Arousal: Physiological And Cognitive Aspects. *Advances in Behaviour Research and Therapy, 8*(3), 101-137.

Wright, L. W., & Adams, H. E. (1994). Assessment of Sexual Preference Using a Choice Reaction Time Task. *Journal of Psychopathology and Behavioral Assessment, 16,* 221-231.

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.

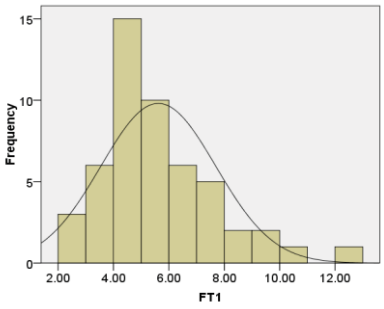
Zamansky, H. S. (1956). A technique for measuring homosexual tendencies. *Journal of Personality, 24,* 436–448.

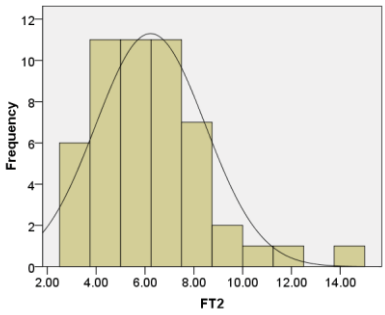
Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin, 75,* 297–329.

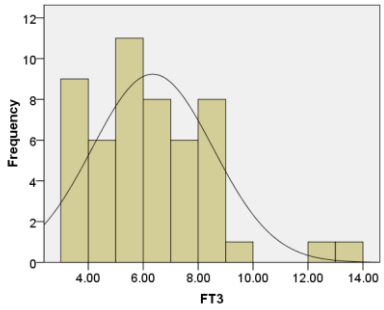
APPENDICES

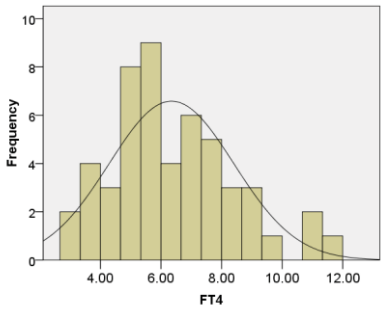
Appendix A – Histograms for Collapsed Tanner Categories

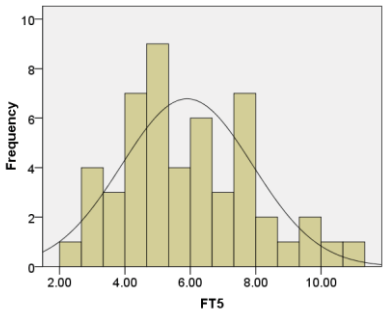
Viewing Time (seconds), N=51

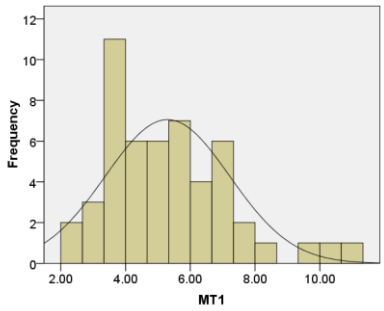


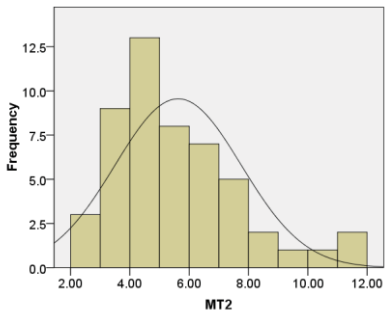


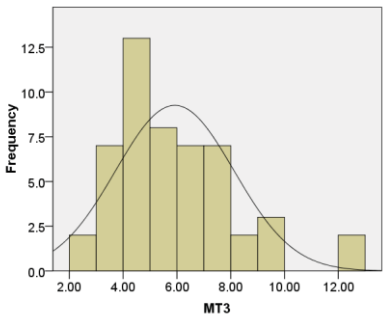


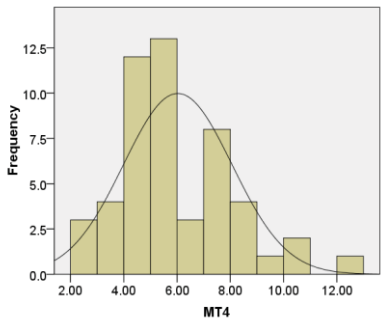


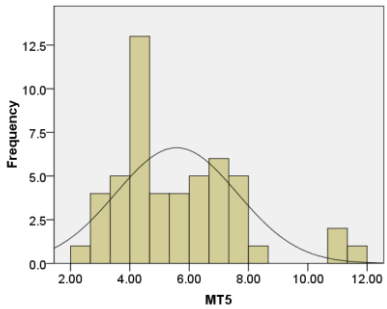


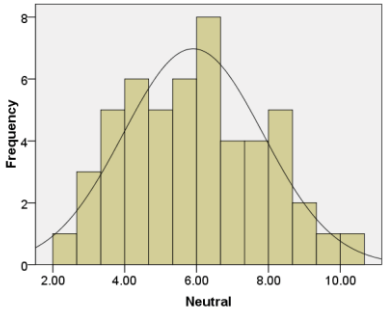




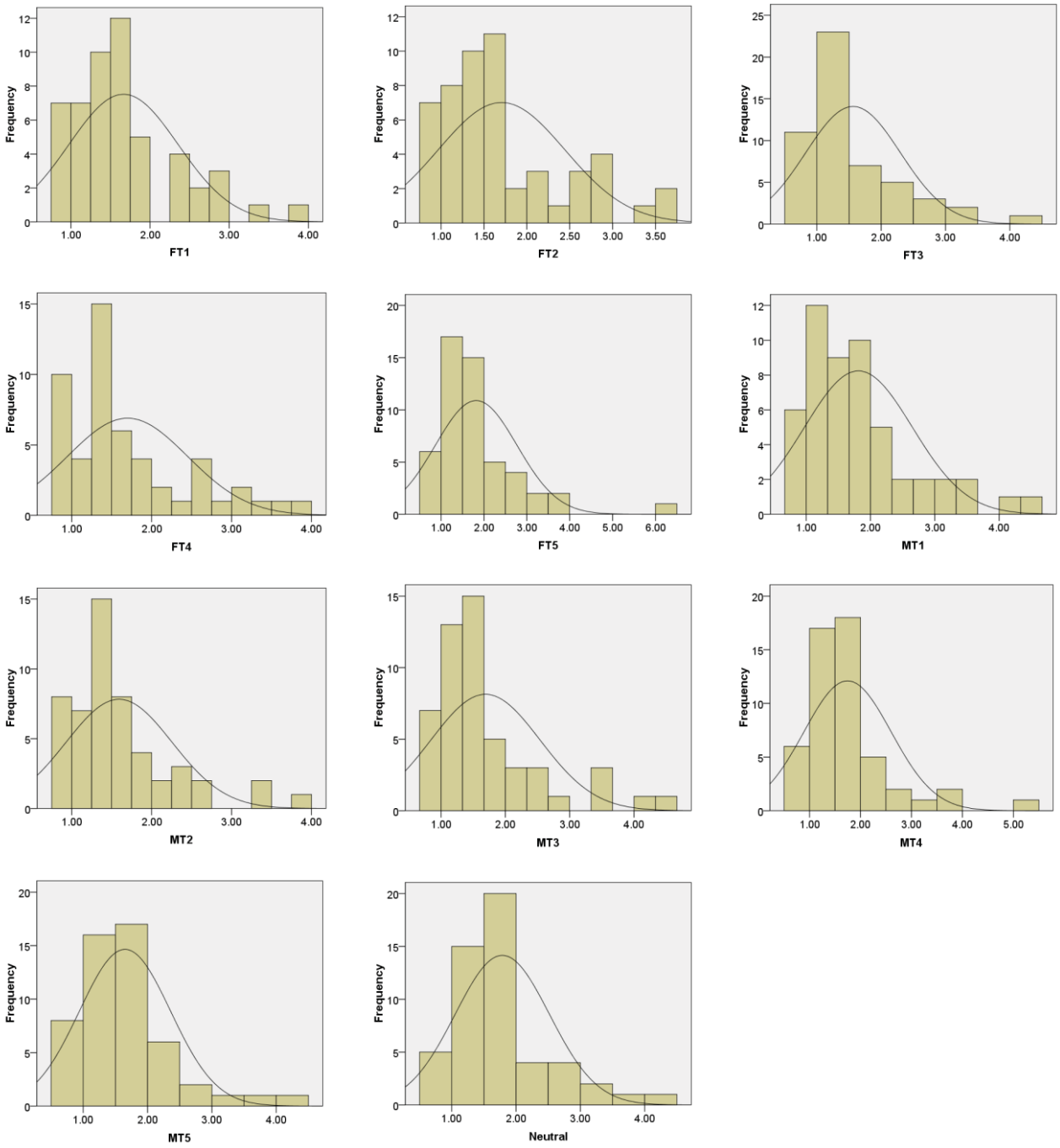








Choice Reaction Time (seconds), N=52



Appendix B – Correlations between and within VT and CRT

Pearson Product-Moment Correlations between VT collapsed categories.

		FT1	FT2	FT3	FT4	FT5	MT1	MT2	MT3	MT4	MT5	Neutral
FT1	Pearson Correlation	1	.940**	.928**	.896**	.791**	.913**	.928**	.927**	.869**	.920**	.738**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
FT2	Pearson Correlation	.940**	1	.971**	.915**	.781**	.919**	.916**	.905**	.866**	.934**	.730**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
FT3	Pearson Correlation	.928**	.971**	1	.935**	.821**	.899**	.900**	.894**	.881**	.923**	.733**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
FT4	Pearson Correlation	.896**	.915**	.935**	1	.897**	.892**	.909**	.906**	.926**	.914**	.735**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
FT5	Pearson Correlation	.791**	.781**	.821**	.897**	1	.743**	.781**	.777**	.832**	.848**	.710**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
MT1	Pearson Correlation	.913**	.919**	.899**	.892**	.743**	1	.952**	.954**	.899**	.903**	.716**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
MT2	Pearson Correlation	.928**	.916**	.900**	.909**	.781**	.952**	1	.974**	.939**	.939**	.760**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
MT3	Pearson Correlation	.927**	.905**	.894**	.906**	.777**	.954**	.974**	1	.941**	.936**	.727**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
MT4	Pearson Correlation	.869**	.866**	.881**	.926**	.832**	.899**	.939**	.941**	1	.933**	.758**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
MT5	Pearson Correlation	.920**	.934**	.923**	.914**	.848**	.903**	.939**	.936**	.933**	1	.771**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		
Neutral	Pearson Correlation	.738**	.730**	.733**	.735**	.710**	.716**	.760**	.727**	.758**	.771**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	

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

Pearson Product-Moment Correlations between CRT collapsed categories.

		FT1	FT2	FT3	FT4	FT5	MT1	MT2	MT3	MT4	MT5	Neutral
FT1	Pearson Correlation	1	.965**	.940**	.965**	.843**	.945**	.958**	.951**	.950**	.953**	.930**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
FT2	Pearson Correlation	.965**	1	.973**	.976**	.909**	.943**	.903**	.910**	.869**	.878**	.866**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000
FT3	Pearson Correlation	.940**	.973**	1	.965**	.940**	.953**	.871**	.900**	.837**	.841**	.840**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
FT4	Pearson Correlation	.965**	.976**	.965**	1	.903**	.951**	.918**	.933**	.899**	.888**	.876**
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
FT5	Pearson Correlation	.843**	.909**	.940**	.903**	1	.908**	.766**	.817**	.736**	.711**	.713**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000
MT1	Pearson Correlation	.945**	.943**	.953**	.951**	.908**	1	.924**	.955**	.895**	.878**	.869**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
MT2	Pearson Correlation	.958**	.903**	.871**	.918**	.766**	.924**	1	.966**	.964**	.967**	.929**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000
MT3	Pearson Correlation	.951**	.910**	.900**	.933**	.817**	.955**	.966**	1	.958**	.931**	.891**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.000	.000	.000
MT4	Pearson Correlation	.950**	.869**	.837**	.899**	.736**	.895**	.964**	.958**	1	.962**	.906**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
MT5	Pearson Correlation	.953**	.878**	.841**	.888**	.711**	.878**	.967**	.931**	.962**	1	.954**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
Neutral	Pearson Correlation	.930**	.866**	.840**	.876**	.713**	.869**	.929**	.891**	.906**	.954**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	

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

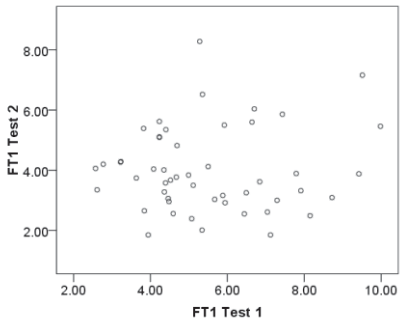
Pearson Product-Moment Correlations between VT and CRT collapsed categories.

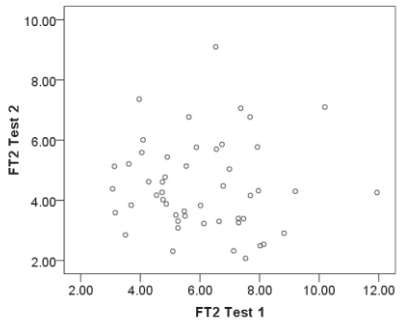
		VT FT1	VT FT2	VT FT3	VT FT4	VT FT5	VT MT1	VT MT2	VT MT3	VT MT4	VT MT5	VT Neutral
CRT FT1	Pearson Correlation	.330*	.288*	.311*	.398**	.384**	.372**	.425**	.430**	.481**	.399**	.302*
	Sig. (2-tailed)	.018	.041	.026	.004	.005	.007	.002	.002	.000	.004	.031
CRT FT2	Pearson Correlation	.373**	.324*	.363**	.438**	.411**	.388**	.442**	.451**	.509**	.414**	.289*
	Sig. (2-tailed)	.007	.020	.009	.001	.003	.005	.001	.001	.000	.003	.039
CRT FT3	Pearson Correlation	.308*	.262	.310*	.407**	.392**	.326*	.385**	.385**	.475**	.359**	.268
	Sig. (2-tailed)	.028	.063	.027	.003	.004	.019	.005	.005	.000	.010	.057
CRT FT4	Pearson Correlation	.303*	.248	.287*	.377**	.364**	.324*	.381**	.393**	.461**	.363**	.271
	Sig. (2-tailed)	.031	.080	.041	.006	.009	.020	.006	.004	.001	.009	.054
CRT FT5	Pearson Correlation	.275	.235	.309*	.397**	.388**	.279*	.358**	.356*	.459**	.314*	.258
	Sig. (2-tailed)	.051	.096	.028	.004	.005	.047	.010	.010	.001	.025	.067
CRT MT1	Pearson Correlation	.339*	.286*	.337*	.427**	.421**	.363**	.430**	.445**	.536**	.402**	.354*
	Sig. (2-tailed)	.015	.042	.016	.002	.002	.009	.002	.001	.000	.003	.011
CRT MT2	Pearson Correlation	.394**	.360**	.378**	.446**	.413**	.433**	.479**	.496**	.532**	.471**	.367**
	Sig. (2-tailed)	.004	.009	.006	.001	.003	.002	.000	.000	.000	.000	.008
CRT MT3	Pearson Correlation	.365**	.321*	.344*	.440**	.412**	.424**	.475**	.497**	.564**	.456**	.350*
	Sig. (2-tailed)	.008	.022	.013	.001	.003	.002	.000	.000	.000	.001	.012
CRT MT4	Pearson Correlation	.307*	.258	.273	.374**	.345*	.378**	.420**	.446**	.478**	.400**	.295*
	Sig. (2-tailed)	.029	.068	.053	.007	.013	.006	.002	.001	.000	.004	.036
CRT MT5	Pearson Correlation	.391**	.348*	.353*	.418**	.403**	.421**	.478**	.484**	.508**	.479**	.368**
	Sig. (2-tailed)	.005	.012	.011	.002	.003	.002	.000	.000	.000	.000	.008
CRT Neutral	Pearson Correlation	.262	.216	.226	.324*	.348*	.301*	.364**	.368**	.417**	.373**	.317*
	Sig. (2-tailed)	.063	.127	.112	.020	.012	.032	.009	.008	.002	.007	.023

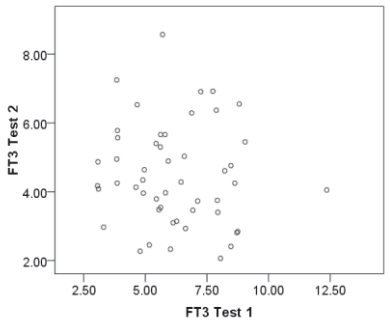
** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

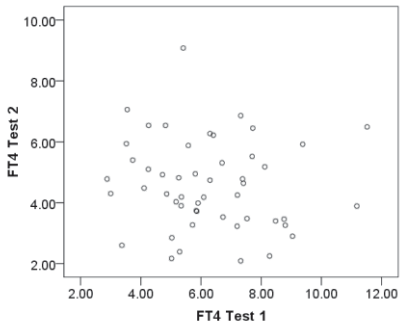
Appendix C – Scatter Plots for Test-Retest

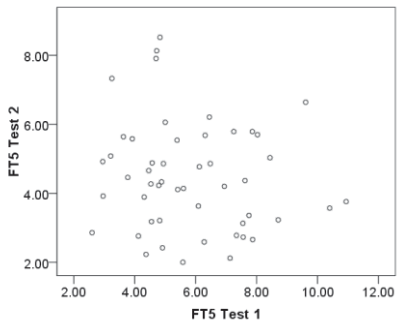
Viewing Time (seconds), N=50

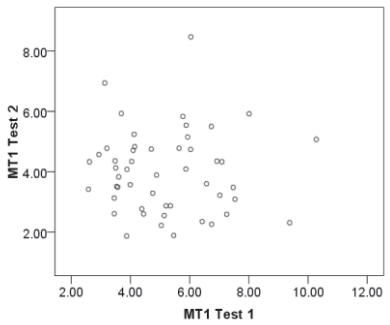


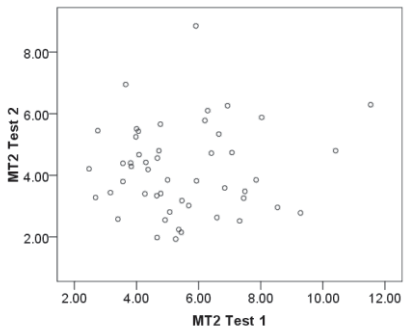


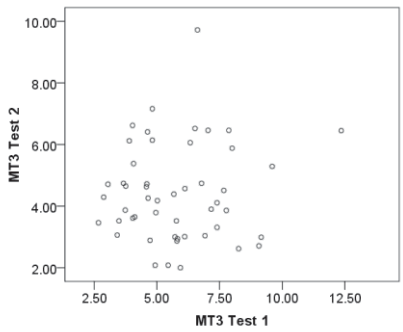


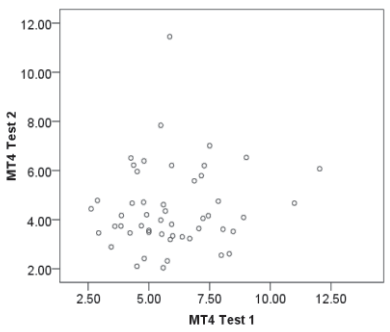


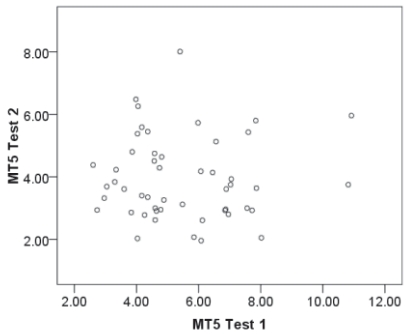


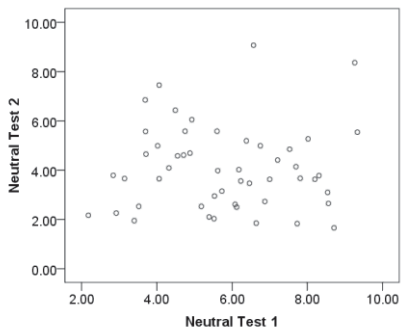




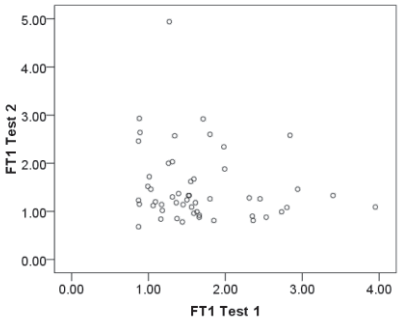


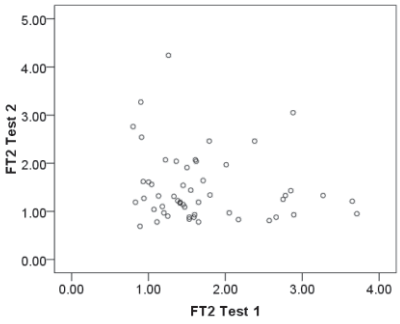


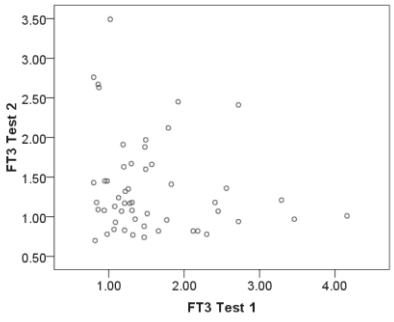


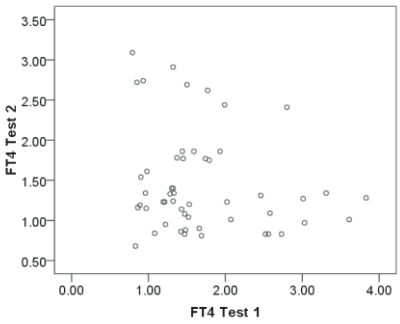


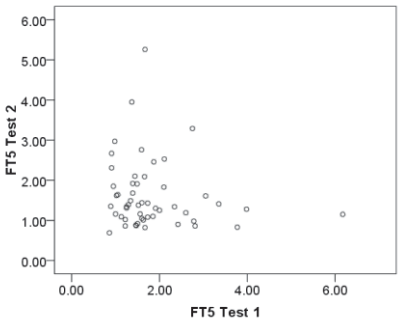
Choice Reaction Time (seconds), N=52

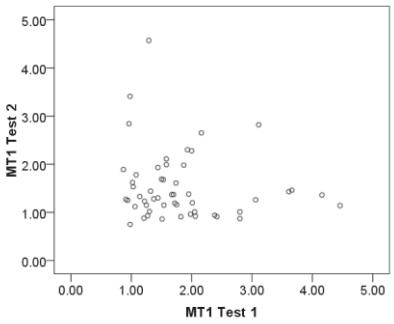


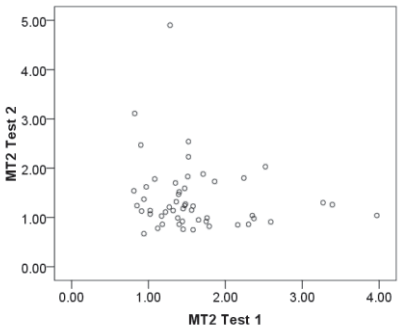


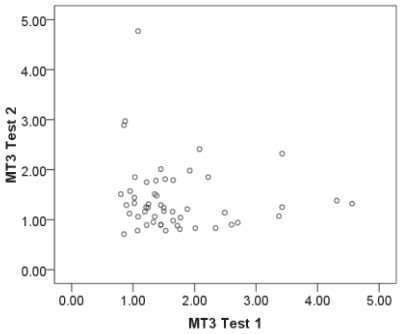


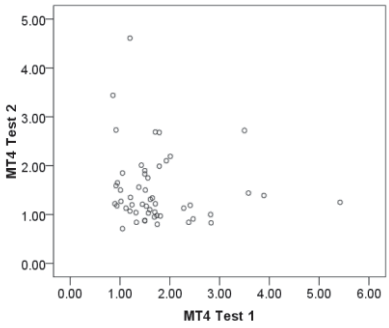


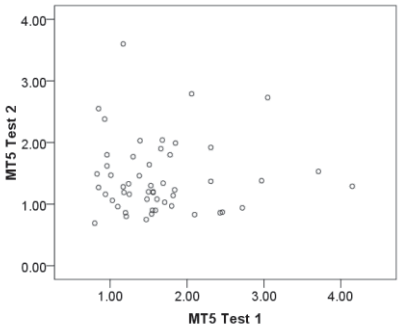


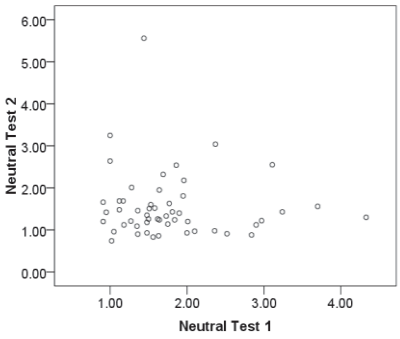












Appendix D - Test-Retest Statistics.

Pearson Product-Moment Correlations between Test 1 and Test 2

	VT (N=50)		**CRT** (N=52)	
	Pearson Correlation	Sig. (2-tailed)	Pearson Correlation	Sig. (2-tailed)
FT1	0.073	0.612	-0.156	0.269
FT2	-0.022	0.878	-0.145	0.305
FT3	-0.088	0.544	-0.175	0.214
FT4	-0.097	0.502	-0.197	0.161
FT5	-0.135	0.350	-0.138	0.328
MT1	-0.021	0.883	-0.126	0.373
MT2	0.063	0.663	-0.157	0.266
MT3	0.083	0.566	-0.115	0.419
MT4	0.113	0.433	-0.109	0.440
MT5	0.006	0.966	0.005	0.973
Neutral	0.026	0.860	-0.070	0.622

VT Wilcoxon Signed Ranks Descriptive Statistics (N=50)

	Mean (seconds)	SD	Min. (seconds)	Max. (seconds)	Percentiles		
					25th	50th (Median)	75th
FT1 Test 1	5.497	1.824	2.570	9.980	4.230	5.090	6.735
FT2 Test 1	6.055	1.905	3.070	11.940	4.738	5.755	7.393
FT3 Test 1	6.193	1.943	3.060	12.360	4.873	5.875	7.890
FT4 Test 1	6.251	1.959	2.880	11.520	4.980	5.880	7.440
FT5 Test 1	5.830	1.957	2.600	10.940	4.513	5.400	7.393
MT1 Test 1	5.175	1.739	2.580	10.280	3.668	4.960	6.458
MT2 Test 1	5.502	1.953	2.470	11.540	4.045	5.035	6.698
MT3 Test 1	5.779	1.973	2.670	12.350	4.108	5.700	7.073
MT4 Test 1	5.946	1.977	2.620	12.030	4.518	5.640	7.243
MT5 Test 1	5.457	1.885	2.600	10.920	4.045	4.795	6.875
Neutral Test 1	5.813	1.853	2.180	9.330	4.255	5.675	7.290
FT1 Test 2	3.993	1.388	1.850	8.280	3.023	3.755	5.095
FT2 Test 2	4.468	1.498	2.070	9.100	3.370	4.265	5.478
FT3 Test 2	4.478	1.465	2.060	8.570	3.445	4.250	5.480
FT4 Test 2	4.585	1.473	2.090	9.080	3.475	4.390	5.610
FT5 Test 2	4.474	1.576	2.000	8.520	3.203	4.300	5.595
MT1 Test 2	4.001	1.351	1.870	8.470	2.870	3.985	4.780
MT2 Test 2	4.176	1.430	1.930	8.850	3.140	4.020	5.273
MT3 Test 2	4.419	1.567	2.000	9.720	3.055	4.220	5.505
MT4 Test 2	4.451	1.701	2.040	11.450	3.448	4.070	5.633
MT5 Test 2	3.936	1.317	1.960	8.010	2.938	3.665	4.763
Neutral Test 2	4.088	1.693	1.660	9.070	2.640	3.785	5.040

CRT Wilcoxon Signed Ranks Descriptive Statistics (N=52)

	Mean (seconds)	SD	Min. (seconds)	Max. (seconds)	Percentiles		
					25th	50th (Median)	75th
FT1 Test 1	1.666	0.690	0.870	3.950	1.173	1.525	1.948
FT2 Test 1	1.705	0.741	0.800	3.710	1.185	1.515	2.040
FT3 Test 1	1.562	0.737	0.800	4.160	1.073	1.310	1.820
FT4 Test 1	1.698	0.752	0.790	3.830	1.213	1.470	2.013
FT5 Test 1	1.819	0.952	0.860	6.170	1.250	1.575	2.075
MT1 Test 1	1.811	0.839	0.870	4.460	1.228	1.580	2.040
MT2 Test 1	1.593	0.663	0.810	3.970	1.173	1.450	1.758
MT3 Test 1	1.691	0.849	0.800	4.560	1.108	1.450	1.910
MT4 Test 1	1.745	0.857	0.860	5.420	1.203	1.545	1.805
MT5 Test 1	1.649	0.708	0.800	4.150	1.173	1.535	1.835
Neutral Test 1	1.786	0.733	0.910	4.330	1.298	1.625	1.990
FT1 Test 2	1.499	0.772	0.680	4.940	0.998	1.250	1.708
FT2 Test 2	1.489	0.725	0.690	4.240	0.955	1.260	1.843
FT3 Test 2	1.367	0.613	0.700	3.490	0.945	1.175	1.623
FT4 Test 2	1.455	0.623	0.680	3.090	1.010	1.255	1.770
FT5 Test 2	1.611	0.862	0.690	5.260	1.058	1.345	1.895
MT1 Test 2	1.529	0.716	0.750	4.570	1.045	1.315	1.758
MT2 Test 2	1.386	0.707	0.670	4.900	0.958	1.195	1.613
MT3 Test 2	1.420	0.696	0.710	4.770	0.958	1.245	1.705
MT4 Test 2	1.503	0.739	0.710	4.610	1.033	1.235	1.810
MT5 Test 2	1.422	0.592	0.690	3.600	0.985	1.275	1.738
Neutral Test 2	1.552	0.798	0.740	5.560	1.120	1.340	1.683

Wilcoxon Signed Ranks Test Results (N=50)

	VT		CRT	
Pair	Z	Asymp. Sig. (2-tailed)	Z	Asymp. Sig. (2-tailed)
FT1 Test 1 – FT1 Test 2	-3.929	.000	-1.498	.134
FT2 Test 1 - FT2 Test 2	-3.934	.000	-1.544	.123
FT3 Test 1 - FT3 Test 2	-4.064	.000	-1.311	.190
FT4 Test 1 - FT4 Test 2	-3.905	.000	-1.457	.145
FT5 Test 1 - FT5 Test 2	-3.205	.001	-1.193	.233
MT1 Test 1 - MT1 Test 2	-3.272	.001	-1.753	.080
MT2 Test 1 - MT2 Test 2	-3.321	.001	-2.031	.042
MT3 Test 1 - MT3 Test 2	-3.417	.001	-1.917	.055
MT4 Test 1 - MT4 Test 2	-3.813	.000	-1.721	.085
MT5 Test 1 - MT5 Test 2	-3.881	.000	-1.867	.062
Neutral Test 1 - Neutral Test 2	-4.108	.000	-1.969	.049

Appendix E – Kruskall-Wallis Statistics

VT Descriptive Statistics for pre-/post-pubescent categories (N=51)

	Mean (seconds)	SD	Min. (seconds)	Max. (seconds)	Percentiles		
					25th	50th (Median)	75th
FT1FT2	5.934	2.132	2.820	13.830	4.530	5.565	7.190
FT4FT5	6.124	1.978	2.915	10.790	4.740	5.710	7.540
MT1MT2	5.461	2.003	2.540	11.595	3.900	5.105	6.465
MT4MT5	5.801	2.009	2.610	11.475	4.415	5.290	7.455

VT Descriptive Statistics for female and male categories (N=51)

	Mean (seconds)	SD	Min. (seconds)	Max. (seconds)	Percentiles		
					25th	50th	75th
Female	6.092	2.023	2.906	12.432	4.710	5.672	7.692
Male	5.689	2.016	2.636	11.538	4.138	5.438	7.066

CRT Descriptive Statistics for pre-/post-pubescent categories (N=52)

	Mean (seconds)	SD	Min. (seconds)	Max. (seconds)	Percentiles		
					25th	50th (Median)	75th
FT1FT2	1.686	.709	.840	3.800	1.179	1.533	1.964
FT4FT5	1.759	.832	.845	4.890	1.223	1.543	2.012
MT1MT2	1.702	.737	.840	4.065	1.203	1.535	1.885
MT4MT5	1.697	.775	.855	4.785	1.188	1.543	1.895

CRT Descriptive Statistics for female and male categories (N=52)

	Mean (seconds)	SD	Min. (seconds)	Max. (seconds)	Percentiles		
					25th	50th (Median)	75th
Female	1.690	.754	0.850	4.090	1.206	1.448	1.898
Male	1.698	.764	0.846	4.452	1.201	1.521	1.898

VT Kruskall-Wallis Tests, Grouping Variable: Sexual Deviance (df = 1, n = 51)

	FT1FT2	FT4FT5	MT1MT2	MT4MT5
Chi-Square	.911	.856	1.190	.277
Asymp. Sig.	.340	.355	.275	.599

CRT Kruskall-Wallis Tests, Grouping Variable: Sexual Deviance (df = 1, n = 52)

	FT1FT2	FT4FT5	MT1MT2	MT4MT5
Chi-Square	.015	.144	.036	.018
Asymp. Sig.	.902	.704	.849	.894

VT Kruskall-Wallis Tests, Grouping Variable: Victim Gender (df = 2, n = 51)

	Female	Male
Chi-Square	1.389	1.322
Asymp. Sig.	.499	.516

CRT Kruskall-Wallis Tests, Grouping Variable: Victim Gender (df = 2, n = 52)

	Female	Male
Chi-Square	.988	.328
Asymp. Sig.	.610	.849

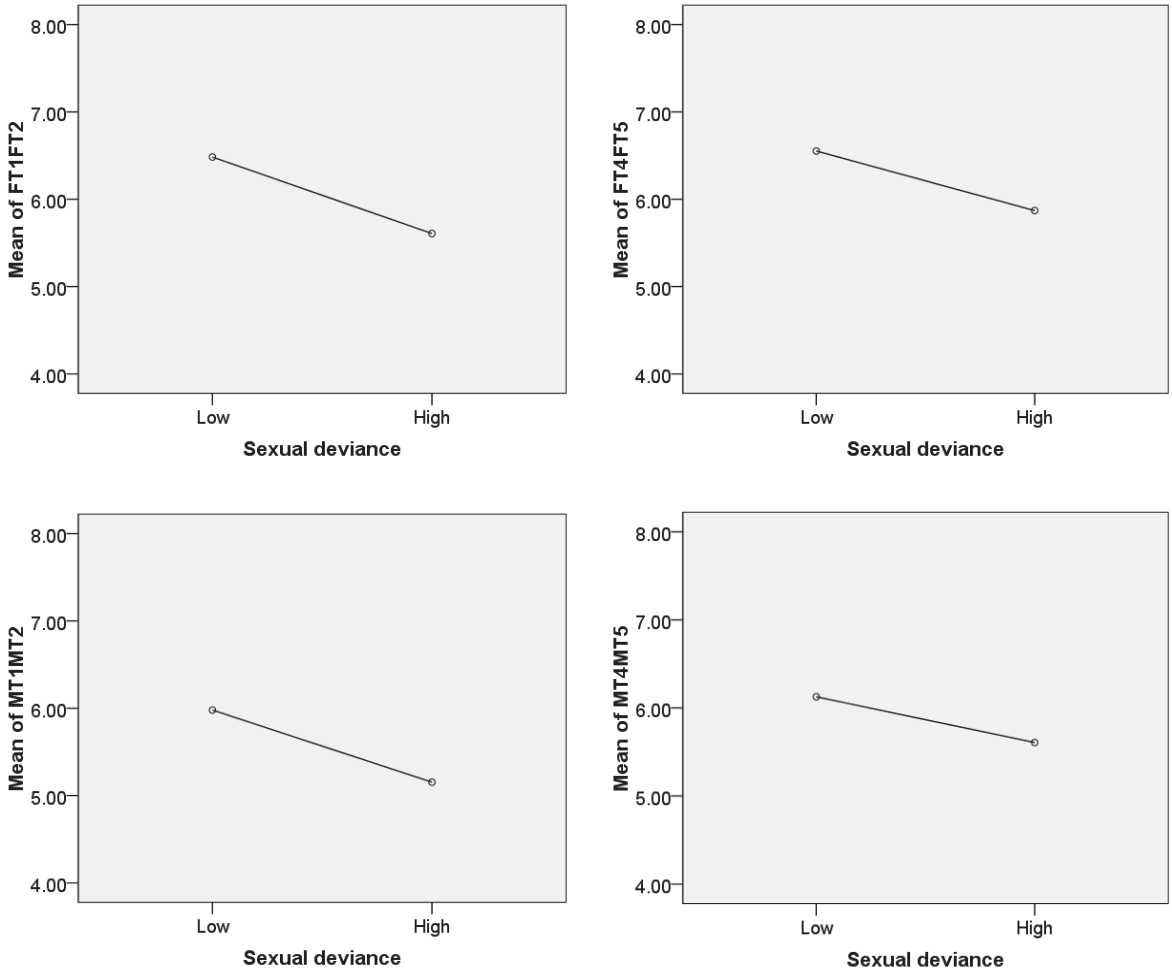
VT Kruskall-Wallis Tests,Grouping Variable: Victim Maturity Category (df = 2, n = 51)

	FT1FT2	FT4FT5	MT1MT2	MT4MT5
Chi-Square	1.201	1.515	.814	1.197
Asymp. Sig.	.549	.469	.666	.550

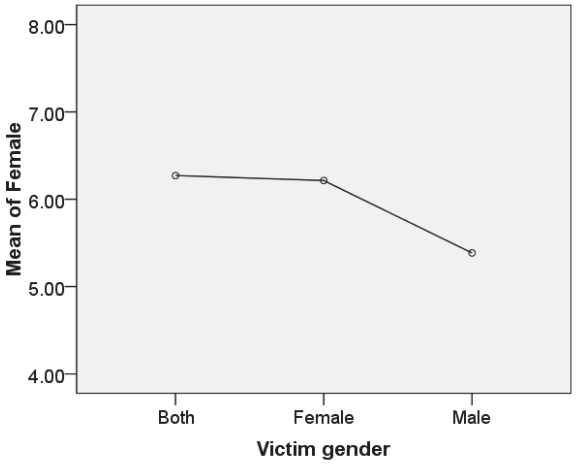
CRT Kruskall-Wallis Tests, Grouping Variable: Victim Maturity Category (df = 2, n = 52)

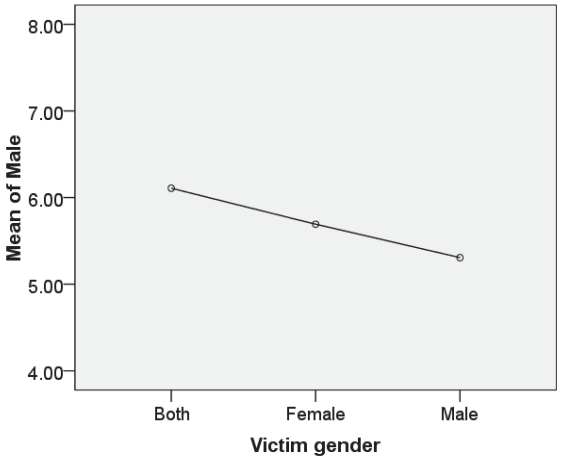
	FT1FT2	FT4FT5	MT1MT2	MT4MT5
Chi-Square	2.201	2.221	2.509	2.124
Asymp. Sig.	.333	.329	.285	.346

VT Means for Sexual Deviance (seconds), N=51

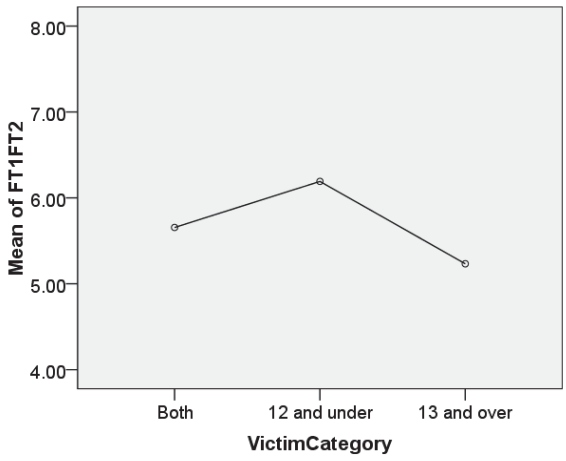


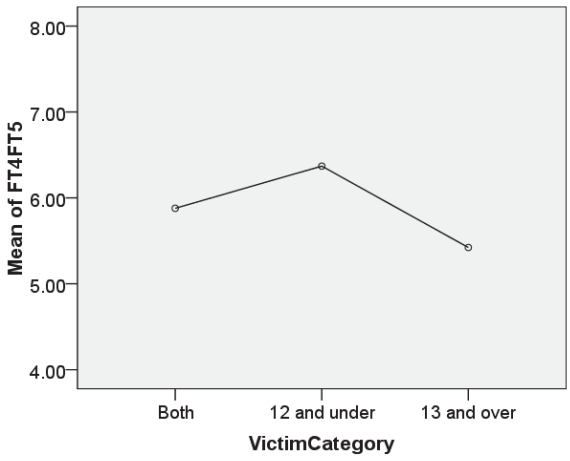
VT Means for Victim Gender (seconds), N=51

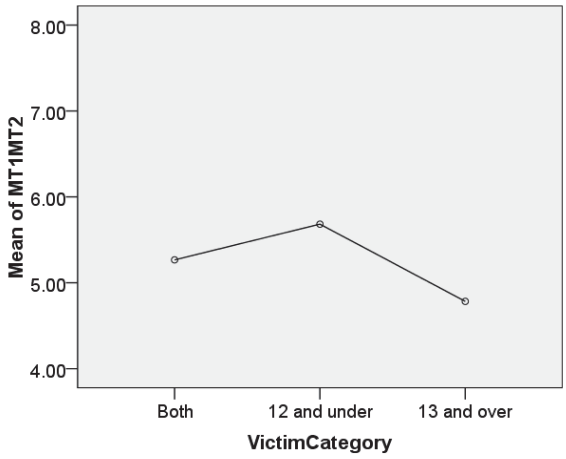


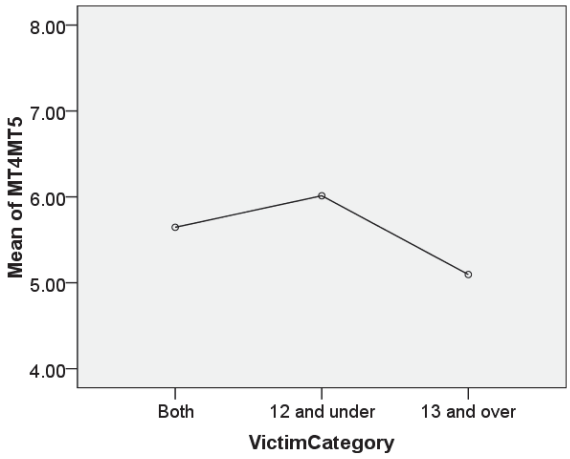


VT Means for Victim Age (seconds), N=51

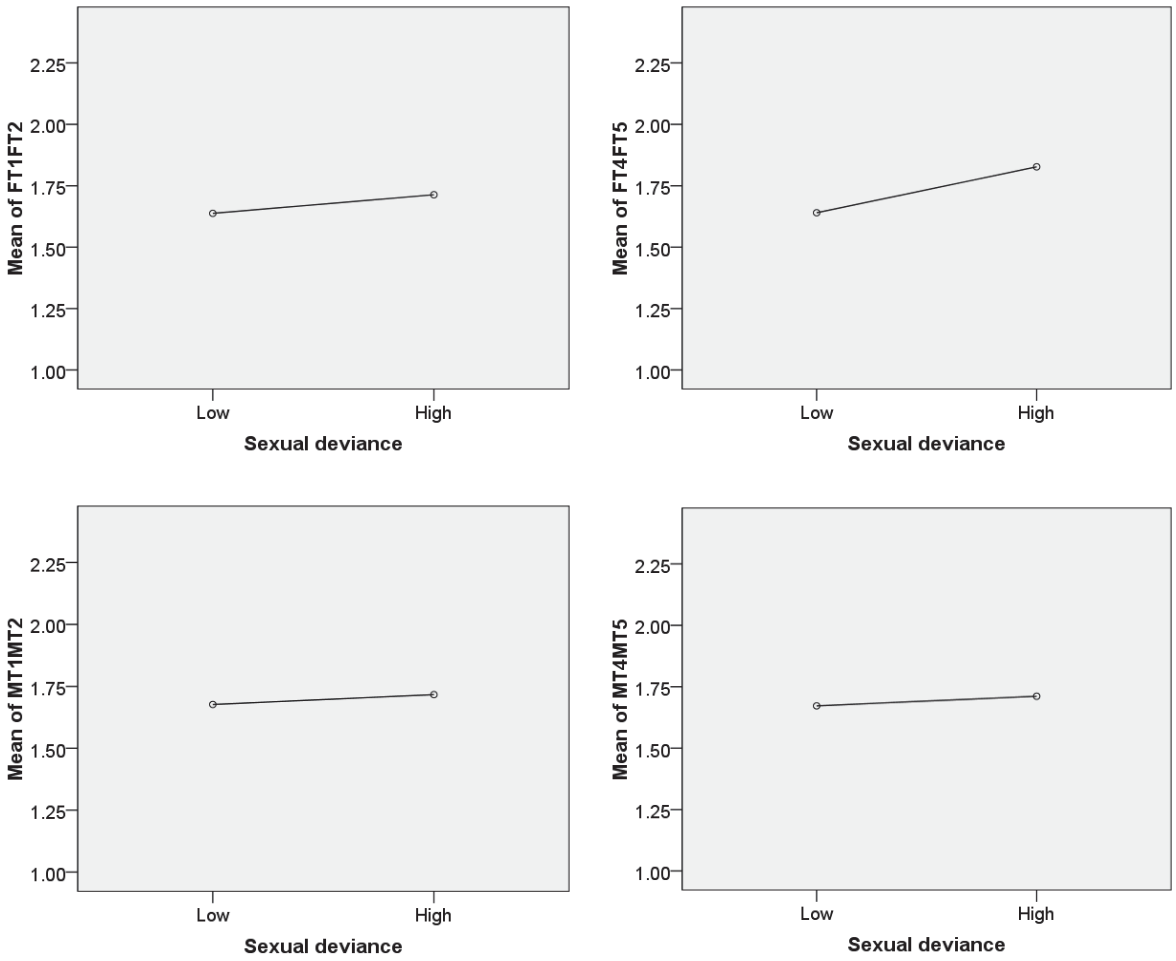




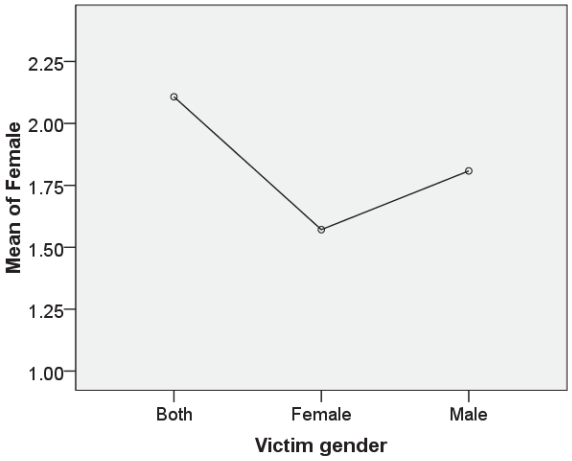


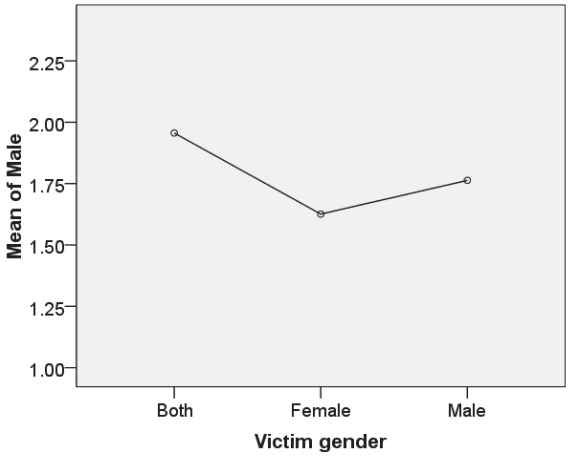


CRT Means for Sexual Deviance (seconds), N=52

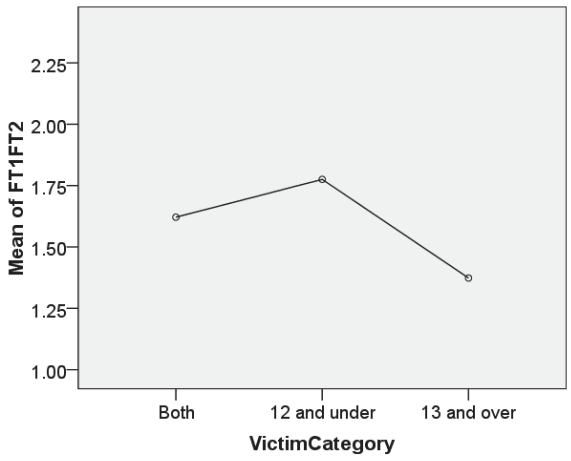


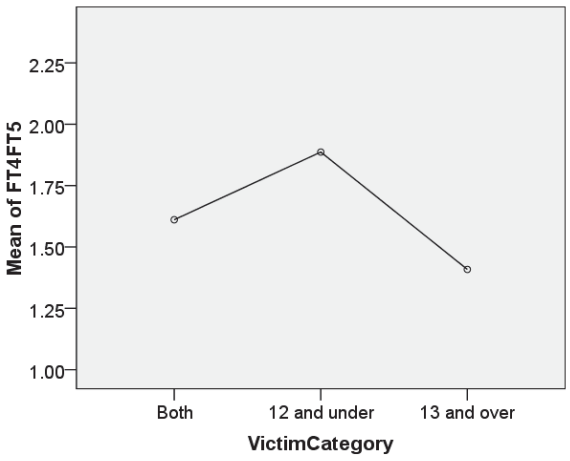
CRT Means for Victim Gender (seconds), N=52

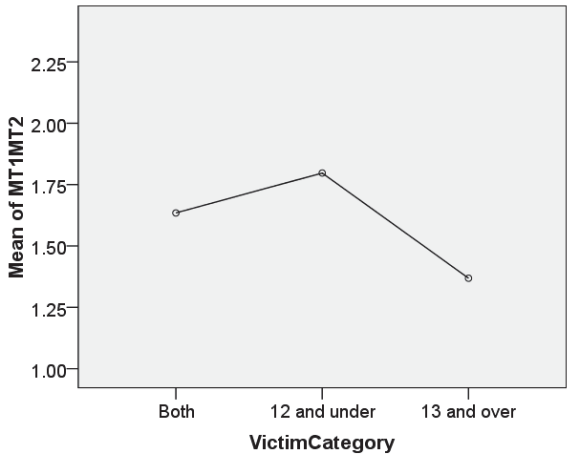


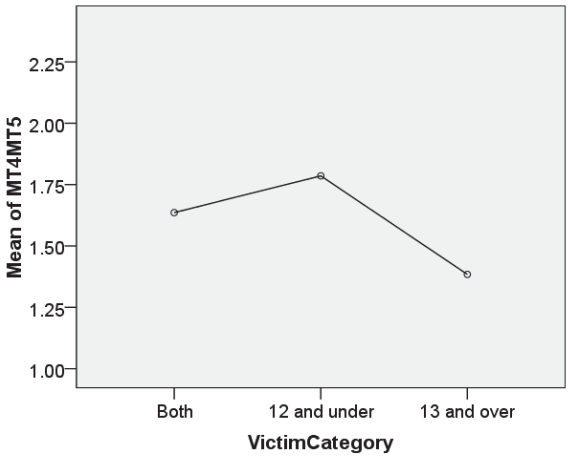


CRT Means for Victim Age (seconds), N=52









Appendix F – Comments to Individual VT/CRT Tests.

Test 1:

1. Worried by mistake making (paused to discuss).
2. *No comments.*
3. Said it was pretty confusing, talked to therapist while assessed.
4. Said he answered they were attractive, but not in sexual way.
5. "Much better than the Plethysmograph, because you can answer what you are attracted to."
6. *No comments.*
7. Lost interest after a while and got bored.
8. He said that he said no to all the questions that asked whether he was sexually attracted to the person
9. *No comments.*
10. *No comments.*
11. "Should I answer with my personal or general opinion?"
12. Has a reading disability. Felt anxious about assessment and talked about it afterwards.
13. Missed some dots, got him thinking, got confused at times.
14. Showed low tolerance to frustration. He responded to the questions at high speed. He said that the purpose was to look only at the dots.
15. Read the questions out loud, said he was getting bored and started singing.
16. "You will see some inconsistencies in the answers, because the girls started plain, but then they became prettier and prettier".
17. Asked whether it was legal to show those pictures, especially the young ones.
18. Was yawning while the assessment was taking place. He would at times lean back on the chair.

19. He noticed a mistake on the example given on the CRT instructions.

20. He had a dark prescription glasses and he stated that at times he had difficulties finding out where the dot was and therefore he guessed that they could be in the centre of the pictures.

21. “It was long”.

22. No comments.

23. No comments.

24. No comments.

25. “Should I answer the questions as I was back then or what I am thinking now. I have done 13-14 months of SAFE. I have a reasonable thought process.”

26. “It takes me a while to read, to process, to compute. I thought it said active and it was attractive and then I realized that. I am not attracted to white people, if it would have been pubescent 14-15 years old, dark skin it would have been different, especially if I am drunk I will be looking at them”

27. Said that he had just woken up when he was called to do the assessment. It was noticed that he was stretching, leaning backwards a couple of times and arranged the chair to lean on it.

28. He was observed leaning backwards. He said "it's getting boring, it's the same pictures". At the end of the VT, said that it had been boring and said "would I be kicked out if I smashed the computer?". "I feel that it repeats the pictures". After the third time he was asked to take time out, and he did it. He then decided to continue with the assessment.

29. "I haven't done anything like this before. After VT practice test, went away to do a Karakia. Said that it was Tapu to see those pictures. Did it the day after, said he was getting fed up. He stated that he had difficulties locating the dot, especially when they were on the centre of the pic.

30. He stated that the dots at times blended with the skin on the CRT

31. “Have never seen pictures like that of kids. I got flashbacks of my own kids. It is about getting used to that.”

32. Said his offending was more of drugs and alcohol. Said he would see how he would react to seeing the pictures. He said that he was not in that frame of mind anymore and that when he saw such pictures on TV, he would change the channel. Talked

loud while answering the questions (VT). Took time to answer the questions, sat back on the chair. Said he liked the neutral pictures. The clouds reminded him of freedom and the water reminded him of diving.

33. No comments.

34. He stated that he was a slow reader.

35. He stated that he avoided looking at pictures, because his counsellor recommended him that. He stated that he avoided masturbation due to being a Jehovah's Witness. He stated that he was computer illiterate. He stated that he had been using thought shifting to avoid focusing on the images sexually. During the VT, he asked one question. He said that he was getting tired and that it repeated itself. During the CRT he read the options aloud. He said he couldn't see two dots.

36. He stated that he was blind on his left eye since age 5, because he feel off a tree. He stated at the end of the assessment that he needed time to process the information, because he had never seen pictures like these before (of children).

37. "I am just scared of screwing it up."

38. No comments.

39. "It took me a while to realize what they were asking, but then I did it."

40. No comments.

41. No comments.

42. Found some faults in the assessment. Said that the dots were easier to locate when they had a black circle around them.

43. No comments.

44. Stated that he got one wrong on the CRT.

45. He stated that it was weird to look at those pictures.

46. Said it was difficult for him to see the pictures of the naked children, but he persevered because he knew it was for his treatment. Said he made some mistakes in the CRT.

47. He stated that he got about 8 wrong on the VT.

48. He asked "how do I rate a person to be attractive, do I do it from my own personal experience?" He stated that he got about 6 wrong on the VT.

49. During the practice test, he stopped, stood up and asked what he should say if he didn't see a male stimulus as attractive or not. He stated that he was not gay. Got 5 wrong on the CRT.

50. No comments.

51. Said he felt suspicious, curious and interested, but was willing to do it.

52. Said that he left gay lifestyle for Christian beliefs, but "deep down" wants a relationship with a male.

Test 2:

1. *No comments.*
2. *No comments.*
3. Wanted to ask questions while taking the test. Asked whether he could go back in a question. Asked to go to the toilet. Said he was confused with the questions "Do you want to be friends with this person?" and "Would you like to have a relationship with this person (non-sexual)?' Not sure about how the question should be answered.
4. *No comments.*
5. *No comments.*
6. *No comments.*
7. Have never seen a naked woman or girl in my life, so this is like pornography for me. "This young good looking fellow, he is about 14. As a Christian, I won't like to have sex with him but if I put that aside I would say yes."
8. *No comments.*
9. *No comments.*
10. *No comments.*
11. *No comments.*
12. *No comments.*
13. *No comments.*
14. *No comments.*
15. *No comments.*
16. Said response option 4 and 5 changed on first practice slide. Said #4 was between 13-14 and now it was 14-16. He said also that option 5 was older than 15 and now it was older than 16.
17. "I got some answers wrong on the dots."
18. *No comments.*

19. While doing the VT assessment, he stopped and asked for the heater to be turned off.

20. He said he had arthritis on his left shoulder and sometimes it made it difficult for him to press the button.

21. No comments.

22. No comments.

23. He expressed concerns about his answer on whether he would like to have sex with a person. He said he would not act on it. He admitted to fantasizing once in a while about children. Said he was attracted to boys 8 to 13. Said he had a co-offender. He would like to have a relationship with him in the future. He stated child porn played a role in his offending. Said he had oral sex with co-offender in jail.

24. No comments.

25. No comments.

26. He stated that even if he was shown an 18 year old white woman, he would not be interested. He stated that he liked darker people. He stated also that he was a slow reader.

27. He was observed stretching and leaning backwards while taking the test. Was observed answering really fast towards the end of the VT, as if he was pressing the same number to change the slides.

28. No comments.

29. Said he was looking only at the top part of the images.

30. No comments.

31. No comments.

32. He said his answer to the CRT items out loud on the first part of the CRT assessment.

33. He stated that on the last assessment he had answered no to being sexually attractive to an older woman and some naked boys. However he stated that he actually felt attracted to them.

34. No comments.

35. Said he didn't know how well he did since he was feeling stressed due to some problems at his work.

36. Read the options of the CRT test out loud.

37. No comments.

38. No comments.

39. He stood up and asked to go to the toilet while doing the CRT, but was asked to stay. Sometimes said his answers out loud when answering questions in the CRT. He stated that he pushed one or two wrong.

40. “During VT my finger pressed a key too hard and some slides went too fast.”

41. No comments.

42. He asked why he had to be assessed two times. Suggested changes to the instructions (Accept button). Said some questions were misspelled in the first sentence on the CRT instructions. Said he got two answers wrong.

43. He took a great amount of time to answer only the first part of assessment. Said at the end that the pictures had triggered some thoughts related to what he talked to his sister, was incoherent on what he meant.

44. No comments.

45. Said it was a shock to the system to see the pictures and to answer the questions.

46. Said he got 4 wrong on the CRT.

47. No comments.

48. No comments.

49. He stated that in the first assessment he held down a key too long and jumped one slide. He was observed stretching and sitting back at times while doing VT. Said he did 4 wrong on CRT.

50. No comments.

51. No comments.

52. *No comments.*

Appendix G – Sample of Te Piriti STU Demographic Questionnaire.

- **Age of offender:**

  ☐ 18 – 25
  ☐ 26 – 35
  ☐ 36 – 45
  ☐ 46 – 55
  ☐ 56 +

- **Current Offences:**

- **Length of Current Sentence:**

- **Number of female children the offender sexually offended against:**

- **Number of male children the offender sexually offended against:**

- **Ages of the offender's female victims of sexual abuse:**

- **Ages of the offender's male victims of sexual abuse:**

- **Relationship of offender to his child victims of sexual offences:**

Marital status:

☐ Married
☐ Single/no current relationship
☐ De Facto/ long term relationship

Ethnic group the offender belongs:

☐ New Zealand/ European/Pakeha descent
☐ New Zealander/Maori
☐ Pacific Islander
☐ Asian
☐ Other

Offender's education level:

☐ Primary or Intermediate school
☐ Some secondary schooling (3rd or 4th form) Form 5
☐ School Certificate, University Entrance, Bursary or similar
☐ Trade certificate or Professional certificate or diploma
☐ Other University degree, diploma or certificate
☐ Other post-school qualification

Offender's stated preference of sexual partner as obtained in questionnaire completed during Plethysmograph assessment:

☐ Children
0 - 5 years old ☐
6 - 10 years old ☐
11 - 13 years old ☐
13 - 16 years old ☐

☐ Adults

Appendix H – STU Consent Form

SPECIAL TREATMENT UNITS

CONSENT TO PARTICIPATE IN TREATMENT

I _____, hereby consent on the basis set out below to participate fully in the Special Treatment Unit (STU) programme designed for those at high risk of re-offending (the programme) and operated by the Department of Corrections ("the Department") at ……prison.

About the programme

I understand that:

- The programme takes about 35 - 40 weeks. This includes about six-weeks for pre-treatment assessment and about four-weeks for post-treatment assessment.
- During the programme (if I am accepted following the pre-treatment assessment), I will participate in activities that are intended to help me to change my attitudes and beliefs about criminal behaviour, and to develop skills to assist me to lead an offence-free and more rewarding life.
- The programme is designed to give me the knowledge and skills to manage my risk of re-offending but will not "cure" me or change my assigned category for risk of re-offending (in so far as this risk is based on factors

that are historical, such as my age, number of convictions and age at first conviction).

- Treatment will mainly be delivered in groups, but I will also have to attend individual treatment, and other treatment groups or meetings and I will be expected to engage in other programme activities. These include, but are not limited to:
- case management (in which my behaviour is reviewed by key staff)
- community meetings (in which everyone in the STU comes together to discuss important issues for the community)
- other structured day activities such as work within the STU and other work that may be offered through CIE, hobbies, kapa haka etc.
- Outside of group sessions I will have to complete individual project work that is related to my psychological treatment.

Use of Information and Confidentiality

- Information gathered about me during my time at the STU will remain confidential unless disclosure to third parties is authorised or required in accordance with the Privacy Act 1993 or otherwise by law.
- I understand that, while participating in the programme, I will be expected to disclose details of my past offending, as well as other personal information.
- I understand that if I talk about serious past offences of which I have not been convicted, this will be treated in confidence but there may be circumstances in which the Department will pass on information about such offending to relevant authorities (e.g. if someone else is in prison for that offending).
- I understand that I will have the opportunity to access and ask for the correction of information about me collected during the programme in terms of Principle 7 (of the Information Privacy Principles contained in section 6 of

the Privacy Act 1993) and that, if any change requested by me is not made, my written statement about that information will be attached to it if I ask for this to be done.

- I understand that if I am having contact with Regional Forensic Services then the staff of the programme and my Forensic Service psychiatrist may share relevant information about me to assist in my assessment and treatment.
- I understand that programme staff will also consult with Department of Corrections' medical staff about problems that may impact on my ability to attend treatment sessions, and discuss my progress with the programme and unit behaviour issues with Prison Services and Probation staff involved in my sentence management.
- I understand STU therapy staff will write a report or reports on my progress in treatment to the New Zealand Parole Board. Further reports will be sent to the Prison Services and also to the Community Probation Office in my home region when I am released from prison. The psychological report/s will be held in a psychological file and on the department's computerised file system, and only those with authorised access will be able to read the report. I also understand that if I am applying for temporary release or being considered for parole that my therapist will write a report to the Prison Services commenting on my progress in treatment and my identified high-risk situations.
- I understand that to ensure I receive the best support and opportunities to change while in the programme, the STU therapy staff will discuss my progress and challenges with other STU staff on an as-needed basis. This will typically occur within case management meetings, and can include discussions with the Unit Manager, PCO, or my Case Officer.
- I understand the STUs have teaching and research roles. During the course of the programme, there may be occasions when, additional to the STU therapy staff, other people will view or participate in group activities for limited periods of time. This could occur by people sitting in on group, use of

a oneway viewing window (if available) and video equipment. I will be told when this is happening and these people will be subject to the same strict confidentiality requirements as the STU staff.

- I understand it is a condition of the programme that, at any time, group and individual sessions may be videotaped with sound, for the purposes of:
    - supervision and professional development of the facilitators running such sessions
    - monitoring of work done on the programme
    - teaching and research
    - providing feedback to me and other fellow group members.
- The Department will ensure that the storage of recordings of group work will be kept secure and that videotapes will be used only for the purposes set out above. All recordings will be destroyed within one month of the completion of the programme unless specifically agreed with me and other group members.
- To make sure that the best treatment possible is provided, the treatment programme, my participation in it, and my progress after release will be evaluated and may be the subject of research. I understand that the results of any such research will be not be published in a form that could reasonably be expected to identify any individual, so my confidentiality will be maintained.
- I understand I may be asked (but not required) to participate in additional special research projects during my time in treatment.

Expected behaviour

I understand that:

- I must be open and honest in all of my dealings with the STU staff and other programme participants, and comply with all programme requirements. This

includes complying with behaviour contracts and prison unit-based activities related to my treatment and video and sound recording requirements.

- I have the right to leave the programme if I choose, though this may mean I have to transfer to another prison, not necessarily the prison from which I originally came to the programme.
- The Principal Psychologist at the STU may discharge me from the programme if my progress is unsatisfactory, or if my behaviour in the STU is not in keeping with the programme or unit retention criteria.

My signature indicates that I have read, understood and accepted these conditions of entry.

_____	_____
Signed	Date
_____	_____
Psychologist/Therapist	Date