

A SCREENING SCALE FOR PRIMARY HEBEPHILIC INTERESTS

DEVELOPING AND VALIDATING A SCREENING SCALE FOR PRIMARY HEBEPHILIC INTERESTS

By

Martina Elena Faitakis

A Thesis Submitted to  
Saint Mary's University, Halifax, Nova Scotia  
in Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied  
Psychology - Forensic Stream

June 2021, Halifax, Nova Scotia

©Martina Faitakis, 2021

Approved: Dr. Skye Stephens  
Supervisor

Approved: Dr. Marguerite Ternes  
Internal Examiner

Approved: Dr. Michael Seto  
Internal Examiner

Approved: Dr. Jay Healey  
External Examiner

Date: June 22, 2021

**Abstract**

Developing and validating a screening scale for primary hebephilic interests

Martina Elena Faitakis

**Abstract:** This thesis sought to create a behavioural measure to assess for primary hebephilia (i.e., a primary sexual interest in pubescent children between the ages of 11 and 14) in male perpetrators of child sexual abuse, and to examine the measure's ability to predict sexual recidivism. In Study 1, two victim characteristics were associated with primary hebephilia (i.e., boy victims under 15 and offences for child pornography), and a behavioural measure of primary hebephilia was not able to be created. Since a measure was not developed in Study 1, Study 2 instead examined the predictive validity of the SSPI-2 and whether the SSPI-2 contributed additional variance beyond the Static-99R in predicting sexual recidivism. Results revealed that total SSPI-2 scores were positively associated with an increased likelihood of sexual recidivism. Challenges of assessing hebephilia and the implications of the association between pedohebephilia and sexual recidivism are discussed.

June 22, 2021

### **Acknowledgements**

There are many people I owe thanks to who have directly and indirectly helped me complete this thesis. First of course, is my supervisor, Dr. Skye Stephens. Skye, thank you for your endless support throughout this degree. Your guidance and patience with me as I took on this steep learning curve was invaluable and I'm hopeful that we will continue to collaborate as colleagues. Although this last year is not how I pictured my degree ending, we have navigated the many challenges of Zoom meetings!

To my committee, who has played a large part in editing and reviewing my thesis. Dr. Michael Seto, who's expertise has far surpassed my expectations. Thank you for forcing me to think deeper about the bigger picture of this thesis and giving me feedback that has made me a better researcher and academic writer. Thank you to Dr. Meg Ternes, who gave me valuable feedback and consistently reminded me to take time for myself and always enjoy the smaller things (like paint by numbers). Thank you, Meg, for also taking me on as a practicum student and allowing me the opportunity to navigate our program evaluation. Thank you to Dr. Jay Healey for agreeing to serve as my external and asking tough yet thought provoking questions during my defense.

I would also like to mention two important people who have been in my corner since I was an undergraduate student who I never got to thank in an Honours thesis. Drs. Sandy Jung & Russ Powell have been instrumental in my success as a researcher and academic, and hopefully will continue to be so throughout my next academic endeavor. Sandy and Russ, I would not be where I am today without your constant encouragement and support, your interests in my research and my life, and your advice and feedback. Between coffee dates at Wilfred's and attending my thesis proposal over Zoom I am

grateful for the incredible mentorship you two have provided me over the years. I am very lucky to have had such great leaders so early in my academic journey.

Thank you to my cohort because we have been through a lot together which includes but is not limited to hurricanes, power outages, online classes, conferences, a wedding, and a global pandemic. To my family and friends who I love so dearly, who show interest in my research and who continue to cheer me on from across the country. I would also like to give thanks to the funding bodies who supported this research over the last two years.

Finally, to Nova Scotia. This province is filled with caring, kind people and beautiful nature. Living five minutes away from the ocean was a privilege that many people don't get to experience in their lifetime. Whenever things got overwhelming during this process, I could always count on a walk along the boardwalk and a front row seat by the ocean to clear my head and ground myself. Two years was not long enough to soak up all that NS has to offer but I'm grateful to have had the opportunity to experience numerous beach trips, weekend getaways, and many, many hikes.

**Table of Contents**

Abstract .....	i
Acknowledgements .....	ii
Chapter 1- General Introduction .....	1
Conceptualization of Hebephilia.....	5
Assessment Tools.....	8
Chapter 2.....	14
Study 1: Creating and validating a behavioural measure of primary hebephilia .....	14
SSPI-2 as a Measure of Pedohebephilia .....	15
Items to Classify Primary Hebephilia .....	16
Present Study.....	19
Method .....	19
Sample.....	19
Measures .....	20
Victim Characteristics.....	20
Phallometric Testing.....	21
Procedure & Data Analysis.....	24
Results.....	25
Development Sample .....	25
Classification Accuracy .....	26
Discussion.....	27
Primary Hebephilia and Victim Characteristics.....	27
Creation of a Measure of Primary Hebephilia: Implications .....	30
Limitations .....	33
Bridging Section .....	35

Chapter 3 .....	36
Study 2: Predictive recidivism of the Revised Screening Scale for Pedophilic Interests (SSPI-2).....	36
Predictive Validity of the SSPI and SSPI-2.....	36
Hebephilia and Sexual Recidivism .....	38
Present Study.....	40
Method .....	40
Sample.....	40
Measures .....	41
Recidivism. ....	41
Opportunity to Reoffend. ....	42
SSPI-2. ....	43
Static-99R.....	43
Procedure & Data Analysis.....	44
Results.....	45
Discussion.....	46
Limitations .....	50
Chapter 4- General Discussion .....	52
Implications.....	53
Future Directions.....	55
References.....	58
Table 1.....	75
Table 2.....	76
Table 3.....	77
Table 4.....	78

Table 5.....	79
Table 6.....	80
Table 7.....	81
Appendix A.....	82
Study Information.....	83
Design Plan.....	83
Sampling Plan.....	85
Variables.....	88
Analysis Plan.....	90
Appendix B.....	94
Study Information.....	95
Design Plan.....	95
Sampling Plan.....	96
Variables.....	99
Analysis Plan.....	101

## Chapter 1- General Introduction

Sexual offending against children is a serious public health and safety concern and can have severe consequences for survivors (Townsend, 2013). In Canada in 2016, 8% of violent crimes were sexual offences, and youths under 18 accounted for 30% of sexual assault victims. Among all age categories, police-reported sexual assaults were highest for youths between the ages of 16 and 17 (Allen & McCarthy, 2016). Sexual offences against children, commonly referred to in the literature as child sexual abuse (CSA), can occur in different forms and are most often perpetrated by someone known to the victim (Finkelhor & Shattuck, 2012). True prevalence rates of CSA are difficult to determine for various reasons. First, CSA is often reported retrospectively when survivors age into adulthood, and a large proportion of CSA goes unreported (Murray et al., 2014). Second, meta-analyses of prevalence rates can have discrepancies due to geographical or cultural origins and different methodological issues (e.g., self-reported versus police documented abuse). In one global meta-analysis, Stoltenborgh et al. (2011) found the prevalence of CSA to be 18.0% for females and 7.6% for males, with a combined prevalence rate of 11.8%.

Another meta-analysis by Maniglio (2009) found that survivors of CSA had a higher risk of developing long-term negative physical, emotional, and psychological health problems, including psychotic symptomatology, depression, anxiety, eating disorders, and personality disorders. Furthermore, negative familial circumstances (e.g., maltreatment, parent mental illness, substance abuse) experienced by victims of CSA are known to contribute to further negative life events, including repeated victimization, familial instability, and low socioeconomic status. Finally, CSA can hinder normal social

growth, in that victims may withdraw from social interactions or act out in ways that are harmful to themselves or others (Maltz, 2002).

Given the prevalence rates and significant negative effects of CSA, researchers have spent a considerable amount of time examining risk factors that contribute to sexual offending against children. Seto (2019) developed the motivation-facilitation theory to explain sexual offending, including offending against children. According to Seto (2019), two broad factors are associated with sexual offending. Motivational factors, which are factors that create a desire to offend (e.g., paraphilias) and facilitators, or factors which increase the likelihood that a sexual offence will occur (e.g., self-regulation problems) in the presence of motivational factors. As a result, different risk factors have been found to be associated with sexual offending against children. For example, antisociality (Seto, 2008; 2013) and emotional identification with children (Hanson & Morton-Bourgon, 2005; Konrad et al., 2018) are both facilitators and risk factors associated with sexual offending against children. One of the strongest motivating factors in understanding CSA and predicting recidivism in individuals who offend against children is having a sexual interest in children (Hanson & Morton-Bourgon, 2005).

The focus of this thesis was on individuals who have a sexual interest in children, which is a distinct chronophilia. Chronophilia is a term that refers to sexual interests regarding different age or maturity categories (Money, 1986). Seto (2016) identified seven distinct age interest categories, classifying sexual interests from birth to old age. Five of these chronophilias are important to briefly highlight due to their relevance to this thesis. First, nepiophilia refers to a sexual interest in infants and toddlers, approximately up to the age of two (Tanner Stage 1). Second, pedophilia which refers to a sexual

interest in prepubescent children, approximately between the ages of three and 10 (Tanner Stage 1). For the purpose of this thesis, nepiophilia and pedophilia were not be differentiated, and nepiophilia was captured under the term pedophilia. Pedophilia is arguably the most well researched form of sexual interest in children and is estimated to be present in 1% of the general male population, worldwide (Seto, 2013; 2016). Pedophilia is also the only age-related paraphilia explicitly included as a disorder in the *Diagnostic and Statistical Manual of Mental Disorders* (fifth ed. [DSM-5]; American Psychiatric Association, 2013). Third, hebephilia refers to a sexual interest in pubescent children, approximately between the ages of 11 and 14 (sex characteristics in Tanner Stages 2-3). Fourth, pedohebephilia refers to a combination of both a sexual interest in prepubescent and pubescent children (Blanchard et al., 2009). Hebephilia and pedohebephilia are not explicitly included in the *DSM-5*; however, they can be diagnosed under Other Specified Paraphilic Disorder (American Psychiatric Association, 2013; Seto, 2018). Lastly, teleiophilia refers to a sexual interest in fully mature adults.

For this thesis, sexual interest in children refers to individuals who have pedophilia, hebephilia, or pedohebephilia. Furthermore, it is important to distinguish the difference between having a sexual interest in children and sexually offending against children. Sexual interest in children and sexual offending against children are not synonymous, as sexual interest in children does not always equate to sexual offending behaviour (Beier et al., 2015). As such, individuals can fall within one of three categories in relation to this: a) those who have a sexual interest in children yet *do not* offend against children, b) those who have a sexual interest in children and *do* offend against children (approximately 50-60% of men who offend against children; Seto, 2018), and c) those

who *do not* have a sexual interest in children and offend against children for other reasons (e.g., opportunistic reasons; 40-50%; Seto, 2018). This thesis sought to distinguish between those who have primary hebephilia (i.e., a sexual interest in pubescent children) and offend against children (group b) from those who do not have a sexual interest in children yet offend against them (group c). Two studies were conducted to establish and validate a behavioural screening measure based on victim characteristics to assess primary hebephilia in men who have previously committed sexual offences against children. Primary hebephilia refers to individuals who may have sexual interests in all age categories of children but have a primary sexual interest towards pubescent children (as determined by phallometric testing). The purpose of Study 1 was to create and establish a measure and determine its discriminant validity and other relevant psychometric properties. Study 2 was to determine the predictive validity of the newly developed measure of primary hebephilia (i.e., The Screening Scale for Hebephilic Interests; SSHI).

The development of the SSHI is important for a few reasons. First, it is useful for clinicians who assess individuals who sexually offend against children, as it would enable clinicians to distinguish between those who offend and have sexual interests in children from those who offend yet do not have sexual interests in children. Offenders who are assessed as having primary hebephilia may have a smaller potential victim pool than those who have pedohebephilia, due to their interests being more limited. Therefore, the measure could also be useful for diagnostic specificity. For example, in theory those who score high on the measure and who have hebephilia should be diagnosed with Other Specified Paraphilic Disorder rather than Pedophilic Disorder. When assessing

perpetrators of CSA, this knowledge is also important in determining risk management and treatment options. For example, a risk management implication is that if someone's sexual preference structure is known it may inform restrictions surrounding access to different groups of children.

### **Conceptualization of Hebephilia**

There is an extensive body of literature on hebephilia, with Bernard Glueck (1955) first introducing it as a distinct age category in his research on individuals who were incarcerated for sexual offences. He originally defined the term as sexual activity with adolescents, distinguishing it from pedophilia (Hammer & Glueck, 1955). The term hebephilia has since been used inconsistently and there has been little consensus on how to best define and understand this sexual interest. For example, hebephilia has previously been defined as a sexual interest in postpubescent adolescents (Franklin, 2010).

Currently, most researchers use the term hebephilia to refer to individuals with a sexual interest in pubescent children (Blanchard et al., 2009). Cantor et al. (2004) classified hebephilia as a "middle" or intermediate group in between those with pedophilia and those with teleiophilia. Research suggests that hebephilia may be more common than pedophilia. In a study by Stephens et al. (2019), 17% of offenders had exclusive hebephilia, 3% of offenders had exclusive pedophilia, and 22% had pedohebephilia. Another study examined 75 male individuals who self-reported a sexual interest in children and found that 26.7% of the men had pedophilia, 30.7% had hebephilia, with the majority (42.6%) having pedohebephilia (Tozdan & Briken, 2015). These findings contradict research by Beier et al. (2015) who found that 20.7% of their sample had exclusive pedophilic interests, whereas only 10.8% had exclusive hebephilic

interests. Interestingly, the largest classification of individuals identified as hebe-teleiophilic (31.1%), indicating that some participants with interests in pubescent children also had sexual interests in adults.

All the aforementioned studies found that most of the offenders in their samples had either pedohebephilia or hebe-teleiophilia, which demonstrates the idea that sexual interest is not always limited to one particular age category. This is relevant as past literature has focused on the idea that a sexual interest in children can be either exclusive or non-exclusive (Beier et al., 2009). Exclusivity is also emphasized in the *DSM-5* which contains a specifier for exclusive or non-exclusive pedophilic disorder (American Psychiatric Association, 2013). The concept of exclusivity suggests that an individual may have a sexual interest in one age category and not others. For example, if an individual has an exclusive sexual interest in children, it means they are not sexually interested in adults at all. In their sample of 75 men, Tozdan and Briken (2015) found that exclusive sexual interest in children was less frequent (20%) than non-exclusivity (80%); however, participants who had a higher exclusive interest in children experienced more stability in their interest and were less likely to believe they could change their sexual interest.

Research has also found that exclusivity is less likely in those with hebephilia compared to those with pedophilia and teleiophilia (Beier et al., 2013). One reason for higher rates of non-exclusivity in hebephilia may be explained by the sexual response gradient. The sexual response gradient postulates that individuals may be sexually interested in a range of targets. Though individuals may show a preference for one target group, this interest can extend to other groups that do not differ significantly from their

main interest (Seto et al., 1999; Seto, Stephens, et al., 2017). For example, an individual who has a sexual interest in pubescent children may also have an interest in adults (one age category above) or prepubescent children (one age category below). For those with an interest in the youngest age category (prepubescent children), it is expected that individuals may also have sexual interests in the adjacent category above (pubescent children). In this example, it would be less likely that those with an interest in prepubescent children would have any interest in fully mature adults, which is two categories above their most preferred age group. Research has generally supported the sexual response gradient based on the significant overlap that exists between pedophilia and hebephilia. For example, one study found that most individuals with pedophilia were also found to have arousal patterns suggestive of hebephilia; however, the reverse was not true in that those with hebephilia did not always have pedophilia (Stephens et al., 2019).

In sum, research suggests that it is uncommon for an individual with sexual interests in children to have arousal towards one specific age group without having any arousal towards a similar age group, especially if the groups are close in age and sexual development (Blanchard et al., 2012; Seto, 2016). Ultimately, this makes it difficult to quantify what constitutes an exclusive interest towards a specific age category. For this reason, this thesis focused on the concept of primary hebephilia as opposed to exclusive hebephilia. Primary hebephilia was operationally defined as individuals who had a much greater phallometric response to pubescent children, relative to adults, and had a higher phallometric response to pubescent children than prepubescent children. As previously mentioned, individuals who have primary hebephilia have an interest in children between

the ages of 11 and 14 and in Tanner Stages 2 or 3 of sexual development. It is important to explicitly determine the primary focus of a person's sexual interests because clinicians may be able to offer more targeted treatment, risk management, and ultimately determine which categories of children are at a higher risk of being victimized. Therefore, it is essential that forensic assessment tools can identify age categories of children that perpetrators of CSA have interests in.

### **Assessment Tools**

There are various assessment methods that are used by clinicians to identify whether an individual has a sexual interest in children, which are briefly reviewed below. Assessing perpetrators of CSA typically involves a psychiatric evaluation, a review of police reports/criminal history, corroboration of collateral sources, and an extensive review/assessment of the individual's sexual history and sexual preference(s) (Coric et al., 2005). A key aspect of forensic assessment is the use of different tools to aid in determining whether a sexual interest in children is present. Over the years, various techniques have been used to determine whether an individual has a sexual interest in children and different assessment tools have been created to measure different aspects of sexual interest (e.g., cognitive or physiological). Four commonly used assessment measures include self-report, cognitive assessments, phallometric testing (measuring patterns of sexual arousal to sexual stimuli), and sexual behaviour history. Although there is not one single tool that can determine whether an individual has a sexual interest in children, measures may be applied simultaneously to increase the validity and reliability of the assessment (Babchishin et al., 2014; Banse et al., 2010).

First, self-report measures are simple and common subjective tools that are

frequently used to assess paraphilias, including sexual interest in children and sexual preoccupation (Moss et al., 2018). One component of self-report involves directly asking clients about their sexual interests during interviews. Another component of self-report is the use of established questionnaires, such as the Multiphasic Sex Inventory II (MSI-II; Nichols & Molinder, 2000), to determine if a sexual interest in children is present. The most obvious challenge with self-report is that individuals may deny a sexual interest in children due to possible legal ramifications or other negative consequences that could result from the disclosure (e.g., conditions prohibiting access to children). Furthermore, the stigma associated with having a sexual interest in children is another factor that can discourage individuals from being honest about their sexual interests (Seto, 2009; 2018). Given that self-report relies solely on honest responding, other measures have been developed to address this limitation.

More recently, cognitive assessment measures have become more popular for assessing sexual interests. Cognitive measures are based on the theory that individuals who have a sexual interest in children will have a processing bias for child stimuli. As a result, these individuals may have longer reaction times or will make more errors in response tasks when viewing child related material (Mokros et al., 2013). One example of a cognitive measure is viewing time, which is a non-intrusive method rooted in the idea that reaction time to child stimuli is indicative of sexual interest. To assess for a sexual interest in children, individuals are shown images of people across the lifespan and asked to rate their degree of sexual arousal to the stimuli. The amount of time an individual spends looking at the stimuli is also recorded, the idea being that the longer they attend to the stimulus the more sexually aroused they are by it (Rosenzweig, 1942). The Explicit

and Implicit Sexual Interest Profile (EISIP; Banse et al., 2010) is an assessment battery that includes a viewing time task to determine whether an individual has a sexual interest in children. The EISIP has demonstrated good validity when discriminating perpetrators of CSA from other offenders and non-offenders (Schmidt & Perkins, 2020), and excellent reliability (Banse et al., 2010). A recent meta-analysis by Schmidt et al. (2017) found that viewing time was able to distinguish men who sexually offended against children from men who did not. The study also found that viewing time was significantly correlated to other measures of sexual interest in children including self-report, other cognitive measures (e.g., implicit association tests), phallometric testing, and measures based on victim characteristics.

A third method of assessing sexual interest in children is phallometric testing, which was originally developed by Kurt Freund (1957) to test sexual orientation in military recruits. Freund later adapted the use of phallometric testing to assess pedophilia in men (Freund, 1965). Research suggests that sexual interest can best be interpreted through patterns of sexual arousal (e.g., Rempel & Serafini, 1995). Penile plethysmography (also known as phallometric testing or phallometry) is an objective technique used to assess for a sexual interest in children. Phallometric testing measures sexual arousal through changes in penile circumference (circumferential testing) or blood volume change (volumetric testing) in response to specific stimuli of people across different ages and sex categories. When testing for a sexual interest in children, individuals are shown stimuli of males and females that vary in sexual maturity (e.g., children, adolescents, and adults) and differences in sexual arousal between the stimuli are recorded. Moreover, when individuals who have an exclusive sexual interest in

children are shown images of adults (presumably images they are not aroused by), the relative change in penile circumference from children to adults may be more indicative of sexual interest in children than the change from baseline to children (Chivers et al., 2010; Marshall & Fernandez, 2000).

Phallometric testing has been consistently and reliably used to distinguish men with a sexual interest in children from those who do not have a sexual interest in children. It is often recognized as the “gold standard” of testing as it has good discriminant and predictive validity (Seto & Lalumière, 2001). A recent meta-analysis by McPhail et al. (2019) used 37 samples to compare perpetrators of CSA and various control groups on phallometric tests, and 16 samples to examine the relationship between phallometric testing and sexual reoffending. Results of their meta-analysis indicated that phallometric testing is both a valid indicator of sexual interest in children and a strong predictor of sexual reoffending for individuals with phallometrically assessed pedophilic, hebephilic, and pedohebephilic interests. Furthermore, different presentation modalities yielded greater validity. For example, slide-only stimuli had the strongest validity for pedohebephilic interests ( $d = 0.75$ ) and was significantly better than video stimuli ( $d = -0.08$ ) at discriminating perpetrators of CSA from the control group.

Despite these findings, phallometric testing is not without limitations. First, different phallometric laboratories follow different procedures, which can lead to variations in psychometric properties across labs. Laboratories may use different stimuli, cut-off scores, and overall processes that can affect the validity of the test as well (Laws, 2009; Wilson & Miner, 2016). This makes it difficult to generalize the psychometric properties for phallometric testing to all procedures. Second, phallometric assessment is

invasive, time consuming, and not always feasible or available for clinicians to use (Laws, 2009). A third limitation of phallometric testing is faking, particularly given the high stakes nature of the assessment. Individuals may be inclined to suppress their levels of sexual arousal by not attending to the stimulus, contracting their muscles, or using cognitive strategies to reduce arousal (Marshall & Fernandez, 2002). A final limitation is the concern that sexual arousal is equated to sexual interest towards a particular group. Although arousal may be considered a good indicator that an individual has a sexual interest towards a particular group, it does not necessarily mean that an individual has or will act on this interest (Merdian & Jones, 2011). Prior research has attempted to objectively define and conceptualize sexual arousal. Sexual arousal varies in its definition, but researchers generally agree that there are three components to sexual arousal including: (1) the aesthetic response, (2) the approach response, and (3) the genital response (Feldman & MacCulloch, 1980; Heiman, 1978; Singer, 1984). Some scholars argue that phallometric testing only measures the genital “autonomic-somatic” component of a sexual arousal response (Singer, 1984), and that phallometric testing does not take a holistic approach when measuring sexual arousal. However, to date no other measure has been able to predict sexual recidivism among perpetrators of CSA with the same accuracy as phallometric testing (Hanson & Morton-Bourgon, 2005).

A final method of assessing sexual interest in children is through an individual’s sexual behaviour history, which addresses some of the limitations of phallometric testing. Certain components of sexual offending history are correlated with sexual interest in children and can be objectively assessed through historical file reviews (e.g., more than one child victim; Seto & Lalumière, 2001). Along with self-report, characteristics of past

sexual offences can be used for assessment purposes. One study by Stephens, Cantor, et al. (2017) found that the combination of self-reported sexual interest in children and past sexual behaviour were better predictors of sexual recidivism together than either indicator alone. Continuing along this line of research, this thesis sought to develop and validate a behavioural measure of primary hebephilic interests and to examine its association with sexual recidivism.

## Chapter 2

### Study 1: Creating and validating a behavioural measure of primary hebephilia

The Screening Scale for Pedophilic Interests (SSPI) is a behavioural measure that was developed by Seto and Lalumière (2001) to identify perpetrators of CSA who require more intensive testing to determine the presence of pedophilia (e.g., phallometric testing). The SSPI was created using a sample of 1,113 adult male offenders with child victims under the age of 15 and is composed of four victim items: any boy victims, multiple child victims, child victims under 12 years old, and any extrafamilial child victims. The measure was created to be easily scored based on sexual offending behaviour history, which can come from previous file information, criminal histories, and clinical interviews. Items are scored dichotomously as present (1) or absent (0), except the boy victim item, which is scored as a two if present. The difference in scoring is because the boy victim item had more weight in predicting phallometric responses to children than the other items. Total scores range from zero to five. Results for the SSPI are promising. Research has demonstrated that perpetrators of CSA who scored a five on the measure were found to be over five times more likely than those who scored a zero to exhibit greater penile responses to children than to adults (Seto & Lalumière, 2001).

In 2017, a revised version of the screening measure (the Screening Scale for Pedophilic Interests-2; SSPI-2) was developed based on literature suggesting that offences related to child pornography<sup>1</sup> were correlated with pedophilia (Seto, Stephens, et

---

<sup>1</sup>The term ‘child pornography’ is used throughout the thesis to refer to child sexual exploitation material (CSEM). The author acknowledges that child pornography is an outdated term that does not acknowledge that the material is sexual exploitation. The term ‘child pornography’ is used since this is the way it is referred to in the SSPI and SSPI-2 and it captures the legal definition in the *Canadian Criminal Code* which includes the making, distribution, possession, or access of explicit sexual activity with a person under the age of eighteen.

al., 2017). The two differences between the SSPI and the SSPI-2 include the addition of the child pornography item and equal weighting across items (total scores range from zero to five). With the addition of the new item, the SSPI-2 has been found to classify individuals with a sexual interest in children significantly better than the SSPI. More recently, a study by Renaud (2019) examined SSPI-2 scores alongside clinician diagnosed pedophilia. Results indicated that as SSPI-2 scores increased, so did clinician diagnosis for pedophilia. For example, 76.8% of those with a SSPI-2 score of four or five were also diagnosed with pedophilia, while only 18.2% of those with a SSPI-2 score of zero had a diagnosis of pedophilia. In terms of intraclass correlation coefficients (ICC), the SSPI-2 had high inter-rater reliability amongst individuals with a child victim,  $r(20) = .94$ . Finally, in another study with a sample of 2,416 perpetrators of CSA, the SSPI-2 was found to be associated with sexual recidivism, particularly within five years of an offender's release (Seto, Sandler, & Freeman, 2017).

### **SSPI-2 as a Measure of Pedohebephilia**

Both the SSPI and the SSPI-2 were originally developed to assess pedophilia in individuals with previous sexual offences; however, Stephens et al. (2019) re-examined the original data used to develop the SSPI-2 to analyze its ability to classify pedophilia, hebephilia, and pedohebephilia. The rationale for the study was that of the five items on the SSPI-2, the only item specific to pedophilia was having a child victim under the age of 12, which was also found to be positively correlated with hebephilia in a previous study (Stephens et al., 2018). Results from Stephens et al. (2019) indicated that the SSPI-2 may be a better indicator of pedohebephilia as it distinguished men with pedohebephilia from those without sexual interest in children; however, the measure predicted hebephilia

only just above chance level accuracy. Furthermore, because the sample had a low base rate of individuals who were exclusively pedophilic, the authors could not examine its ability to classify exclusive pedophilia. Results from Stephens et al. (2019) suggest that it is more difficult to separate hebephilia and pedophilia, and that further research is needed to classify individuals as having primary pedophilia or hebephilia. Because of this limitation, the present study focused on developing a measure of primary hebephilia based on past sexual behaviour. Assessing for and identifying those with primary hebephilia is important so that treatment can be targeted to meet individual needs and risk management can be more specific.

### **Items to Classify Primary Hebephilia**

There are different items from the SSPI-2 that may be particularly relevant when assessing primary hebephilia. It was anticipated that two items from the SSPI-2, multiple child victims under 15 and child pornography, would accurately classify those with primary hebephilia. First, having multiple child victims under the age of 15 has been found to be indicative of hebephilia in past studies (Stephens et al., 2018; Stephens, Seto, et al., 2017). An association between multiple child victims under 15 and primary hebephilia was hypothesized because of the considerable overlap that exists between pedophilia and hebephilia (Stephens et al., 2019). Furthermore, an offender who has primary hebephilia may have victims who are prepubescent for other reasons, like victim availability. Multiple child victims under 15 also accounts for the concept of an expanded victim pool, which refers to the idea that perpetrators may have a preferred victim pool; however, this preference can change over time and may expand to other age categories when their preferred victim age is unavailable to offend against (Heil, Ahlmeyer, &

Simons, 2003). Lastly, it is important to note that age is an imperfect proxy of sexual development, so including victims under 15 as a potential item could capture victims who although were younger in age, were pubescent in appearance.

The second item that was anticipated to be associated with primary hebephilia was child pornography offences. Previous research by Seto and Eke (2015) found that self-reported or diagnosed hebephilia in individuals with child pornography offences was associated with sexual recidivism. Similarly, Stephens, Cantor, et al. (2017) found hebephilia to be associated with non-contact sexual recidivism, which included offences for child pornography. Although research indicates that a large proportion of child pornography depicts prepubescent girls (Seto et al., 2018; Westlake, 2020), the *Criminal Code of Canada* includes any person under the age of 18 in their definition of child pornography, which is inclusive of victims who would be pubescent in appearance. Given that child pornography does include depictions of pubescent children, it was worthwhile to examine whether the presence of child pornography offences would accurately classify individuals with hebephilia.

There are also additional items that were thought to improve classification accuracy in a measure of primary hebephilia. Past research has found that indicators of hebephilia, but not pedophilia, were consistently associated with a greater number of victims between the ages of 15 and 16 (Stephens et al., 2018). Although the cut off age for hebephilia based on the Tanner Stages is roughly 14, the sexual response gradient suggests that individuals who have a sexual interest in pubescent children may also exhibit some arousal to the adjacent categories above and below pubescent children. It is therefore reasonable to consider that victims between 15 and 16 could accurately classify

those with primary hebephilic interests. The logic behind including victims between 15 and 16 as a proposed item was also based on the idea of an expanded victim pool as discussed for the first item. Victims between the ages of 15 and 16 could fall into the expanded victim pool for those with primary hebephilia when their preferred victim is not available. Furthermore, victims from this age group may be easier to access because they are more likely to be in the community with less adult or parental supervision than younger children. Because findings from Stephens et al. (2018) found an association between hebephilia (both self-reported and phallometrically assessed) and victims between the ages of 15 and 16, it was justifiable to include this item as a potential correlate of primary hebephilia.

Furthermore, gender polymorphism (i.e., offending against both girls and boys) has also been found to be associated with pedophilia, but likely not hebephilia. Levenson et al. (2008) found that offenders with young victims were more likely to have both boy and girl victims and that those with pedophilic interests were more likely to have either boy victims or victims of any gender. The finding that pedophilia is associated with gender polymorphism is intuitive in that prepubescent boys and girls would have similar physical features. Children at this age have not yet developed secondary sex characteristics, and individuals with pedophilic interests may not differentiate their offending between the two genders. On the other hand, those with hebephilic interests have preferences for children who have begun to develop secondary sex characteristics, meaning they may be more likely to discriminate between girls and boys, offending solely against one gender and not the other. It was expected that an absence of gender

polymorphism or an individual who solely offended against a victim of one gender, would be indicative of hebephilia rather than pedophilia (Blanchard et al., 2012).

### **Present Study**

The current study extended research by Stephens et al. (2019), as the authors did not explicitly examine the performance of individual SSPI-2 items in their ability to classify primary hebephilia. The purpose of the present study was to determine victim characteristics that were associated with primary hebephilia and examine if a scale could be created that could accurately classify individuals with having primary hebephilia. First, it was hypothesized that two items on the SSPI-2, multiple child victims under 15 and child pornography, would be associated with primary hebephilia. Exploratory analyses were also conducted to examine the remaining three SSPI-2 items in their ability to identify primary hebephilia. Second, it was hypothesized that two additional items not from the SSPI-2 (i.e., victims between the ages of 15-16 and offenders who did not exhibit gender polymorphism) were more likely to also be associated with primary hebephilia. Finally, it was hypothesized that the new measure would be able to accurately classify individuals with primarily hebephilic interests and have strong psychometric properties.

### **Method**

#### **Sample**

The study used an archival database from the Sexual Behaviour Clinic (SBC) at the Centre for Addiction and Mental Health (CAMH) in Toronto, Ontario. Individuals were assessed between 1995 and 2011, and the initial dataset included 3,343 cases of men who were referred for an assessment of their sexual interests using phallometric testing.

Men who had no victim information were removed from the dataset which left a total of 2,464 offenders. To create the current sample, only those who had a child victim under the age of 15 ( $n = 1,900$ ) were included.<sup>2</sup> We used this victim age because the same cut-off was used for the development of the SSPI and the SSPI-2 in their samples of individuals who previously offended against children. Furthermore, individuals who were phallometrically assessed to have pedophilia only ( $n = 61$ ), and those who had pedohebephilia but responded more to prepubescent children than pubescent children ( $n = 184$ ) were removed. The final sample size included 1,655 participants and was comprised of those in the comparison group who had no sexual interest in children ( $n = 1,088$ ) and those in the target group who had primary hebephilia ( $n = 567$ ). Descriptive statistics for Study 1 are presented in Table 1. To promote transparency, the research questions, hypotheses, and planned analyses were pre-registered on the Open Science Framework ([osf.io/urbqn](https://osf.io/urbqn)) and are described in Appendix A.

## Measures

**Victim Characteristics.** The database included information on victim characteristics for each offender. This information came from two sources: a) official file information which included previous assessments and official criminal history documentation (e.g., police reports), and b) self-reported sexual history obtained through clinical interviews. When a discrepancy in victim information occurred between official file information and self-reported information, the source which had the highest number of victims was used. Victim information including gender, age, and relationship to the

---

<sup>2</sup> Offenders were only included in the sample if they had at least one child victim under the age of 15. They may have had victims in other age categories over 15 as well but must have had at least one child victim under 15 to be included. This means that those who only had victims who were 15 and older were removed from the sample.

offender was recorded in the original database. All the original SSPI-2 items were examined including: any boy victims under 15, more than one child victim under 15, any child victims under 12, any extrafamilial child victims under 15, and charges or admissions of child pornography offences. Finally, two additional victim characteristics were analyzed including victims between the ages of 15 and 16, and the absence of gender polymorphism. In the original database, gender polymorphism (offending against both males and females) included both child and adult victims. However, since the aim of this study was to develop a measure for individuals who offend sexually against children, the original variable was recoded so that gender polymorphism was only in reference to child victims. All these items were scored dichotomously as present or absent.

**Phallometric Testing.** All participants referred to the clinic underwent phallometric testing during their sexological assessment to inform diagnoses related to sexual interest (e.g., pedophilic disorder). Prior to phallometric testing participants gave their written and verbal consent to participate in the assessment. Participants were subsequently placed in a room with a camera aimed at their face to ensure they were attending to the stimuli. A technician was stationed in a separate room behind the participant with a one-way mirror which enabled them to ensure the participant was attending to the stimuli. Participants were instructed to place a cuff at the base of their penis and a plastic cylinder over the cuff to create an airtight seal. Changes in penile blood volume were recorded and corresponded to changes in air displacement in the cylinder.

Participants were then shown four trials of seven different stimulus sets composed of still images and audio stories based on biological sex (male or female) and sexual

maturity (prepubescent, pubescent, or adult). The sets included: a) prepubescent girls (aged 3 to 11), b) prepubescent boys (aged 5 to 11), c) pubescent girls (aged 12 to 14), d) pubescent boys (aged 12 to 14), e) adult females (aged 20 to 34), and f) adult males (aged 19 to 41). The seventh category of stimuli was a neutral landscape image with an audio description and was presented as a control category. A pediatric endocrinologist was involved in rating and ensuring the visual stimuli resembled the corresponding Tanner stage of sexual development on a scale of one (indicating prepubescent) to five (indicating a full mature adult; Blanchard et al. 2009; Tanner 1978). Participants viewed three images of the stimuli on a large screen: the front and back of the individual, and a close-up of the genitalia. Participants simultaneously listened to audio stories through headphones describing a passive sexual interaction with someone who corresponded to the stimulus in terms of sex and sexual development. The female audio stories were narrated by a woman, and the male audio stories were narrated by a man.

During each trial, penile blood volume was recorded four times per second. Before showing each stimulus, the technician waited for the participant's blood volume level to return to their baseline. The data collected during the procedure was converted to  $z$ -scores for each trial and the scores were averaged for each stimulus category to create seven  $z$ -scores per stimulus set. A hebephilia index was calculated by subtracting the maximum average score to adults from the maximum average score to pubescent children, regardless of sex. The pedophilia index was calculated in the same manner but using prepubescent child stimuli. A positive score indicated a greater relative response to children, while a negative score indicated a greater relative response to adults. A cut off score of  $z > 0.25$  was used to classify the test as indicative of hebephilia or pedophilia.

Based on the hebephilia and pedophilia indices, four groups were created: a) those who met the index score for pedophilia only (pedophilia), b) those who met the index score for hebephilia only (hebephilia), c) those who met the index score for both pedophilia and hebephilia (pedohebephilia), and d) those who did not meet either index score and therefore had no phallometrically assessed sexual interest in children. As previously discussed, the two groups used for the current analyses included those: a) who reached the phallometric cut-off score on the hebephilia index only or reached the phallometric cut-off score for both the pedophilia and hebephilia index, but had a higher score on the hebephilia index, and b) those who did not reach the phallometric cut-off score for either the pedophilia or hebephilia index and were used as the reference group.

The phallometric procedure in the present study has been the focus of many research studies dedicated to establishing its psychometric properties. Sensitivity is the ability for a test to correctly identify the condition when it is present, and specificity is the ability of a test to correctly identify the condition as absent when it is not present. Cantor and McPhail (2015) found moderate sensitivity and high specificity for the CAMH phallometric procedures. Sensitivity was 70.0% for hebephilia, 46.9% for pedophilia, and 75.3% for pedohebephilia. Specificity was 90.7% for hebephilia, 100% for pedophilia, and 95.3% for pedohebephilia. Furthermore, Stephens, Cantor, et al. (2017) found that the use of this phallometric procedure demonstrated predictive validity for sexual non-contact recidivism, but not sexual contact recidivism. Discriminant validity has also been examined for the current phallometric procedure and has been found to discriminate well between perpetrators of CSA and individuals who do not sexually offend against children (e.g., Blanchard et al., 2009; Freund & Blanchard, 1989; Freund & Watson, 1991).

### **Procedure & Data Analysis**

Offenders were referred for a sexological assessment at CAMH which included a clinical interview and phallometric testing. The data was recorded by a technician and entered by a research assistant and an archival database of the assessment information has been maintained at the clinic. Only individuals who provided their consent for their data to be used for research were included in the present study. The archival database used for this study is large and includes several variables that have been used in other research studies. Some examples of variables that were included in the database but were not used for the present study include the offenders age of puberty, the age an offender first had intercourse with a female and/or male, and how many consenting sexual partners an offender has had.

Before analyzing the data, the total sample was randomly divided into a development and a validation sample. To do this, the target and comparison groups were randomly divided into two subsamples and then recombined to create the two samples. First, the target group was divided in half to create two subsamples ( $n = 283$  and  $n = 284$ ) and the comparison group was divided in the same way (both  $n = 544$ ). One subsample from the target and one subsample from the comparison group were combined to create a development sample ( $n = 827$ ) and a validation sample ( $n = 828$ ). The development sample was used for the creation of the measure to address the first three hypotheses, and the validation sample was used to determine the classification accuracy of the new measure.

Chi-square analyses were performed to address the first two hypotheses and exploratory analyses were conducted to determine which individual items (both from the

SSPI-2 and other additional items) were associated with primary hebephilia. Logistic regression was used to address the third hypothesis to determine which items should be retained on the measure. Furthermore, Area under the Curve (AUCs) were also used to determine the ability of the measure to classify those who have primary hebephilia.

AUCs represent the probability that a randomly selected offender with primary hebephilia would score higher on the SSHI than an offender who has no sexual interests in children.

The second sample contained the other half of the participants ( $n = 828$ ) and was used for validation purposes. Validation is important to ensure the measure could accurately classify the remainder of the sample as having primary hebephilia. AUCs were used for classification accuracy in determining the ability of the SSHI to correctly identify individuals with primary hebephilia. Finally, psychometric properties including sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV) were examined by using the original total sample ( $n = 1,655$ ).

## **Results**

### **Development Sample**

Chi-square results for the development sample are presented in Table 2. The two items from the SSPI-2 that were originally hypothesized, multiple child victims under the age of 15, and offences for child pornography, were both positively associated with primary hebephilia. Two other items from the SSPI-2 that were examined as exploratory analyses, any boy victim under 15, and any child victim under 12, were also positively associated with primary hebephilia. The two additional items that were hypothesized, one or more victims between 15 and 16, and the absence of gender polymorphism, were not significantly associated with primary hebephilia. Finally, the fourth SSPI-2 item,

extrafamilial child victims under 15, which was examined as an exploratory analysis, was also not significantly associated with primary hebephilia.

Results of the logistic analysis found that the four-predictor model provided a statistically significant improvement over the constant-only model,  $\chi^2(4,827) = 27.64, p < .001$  (see Table 3). The Nagelkerke pseudo  $R^2$  indicated that the model accounted for 4.5% of the total variance, which suggests that only a small amount of variance was accounted for by the items in the model. The only two items that remained significant in the logistic regression were boy victims under 15, and possession of child pornography. Offenders who had primary hebephilia were 1.43 times more likely to have any boy victim(s) under the age of 15 than those who had no sexual interest in children. Furthermore, offenders who had primary hebephilia were 2.55 times more likely to have charges or convictions for possession of child pornography than those who had no sexual interest in children.

### **Classification Accuracy**

The significant items from the logistic regression were combined to create a two-item measure called the “SSHI.” Using a  $> .25$  criterion for the development sample ( $n = 827$ ), the “SSHI” was found to be just significantly above chance level in its ability to classify individuals as having primary hebephilia,  $AUC = .55, SE = .02, 95\% CI [.51, .59], p = .014$ . For the validation sample ( $n = 828$ ), the ability of the “SSHI” to classify individuals as having primary hebephilia versus having no sexual interest in children was also just above chance level, but not significant,  $AUC = .54, SE = .02, 95\% CI [.495, .579], p = .084$ . Lastly, the SE, SP, PPV, and NPV for the two “SSHI” cut-off scores were calculated for the total sample and are presented in Table 4.

## Discussion

The goal of the present study was to determine victim characteristics associated with primary hebephilia and to combine these items to create a measure; however, a measure was unable to be developed with the proposed victim characteristics. Although most of the hypothesized victim characteristics were positively associated with primary hebephilia (i.e., multiple child victims under the age of 15, offences for child pornography, any boy victim under 15, and any child victim under 12), only two of the items remained significant (i.e., boy victims under 15 and offences for child pornography) when considering all significant victim characteristics in the logistic regression model. While it was not feasible to create a measure with two items, they were nonetheless combined to create a pseudo measure to examine the “SSHI’s” ability to identify offenders with primary hebephilia and those with no sexual interest in children. For both the development and the validation sample, the two-item measure was only able to classify offenders as having primary hebephilia just above chance level, which indicates that the pseudo measure had poor classification accuracy. Taken together, these findings highlight that a measure of primary hebephilia was unable to be created and that the SSPI-2 may currently be the most useful tool for clinicians to use when assessing for pedohebephilia (Stephens et al., 2019).

### Primary Hebephilia and Victim Characteristics

Although a measure of primary hebephilia was not developed, it is important to highlight some of the individual items and their associations with primary hebephilia. Regarding the first hypothesis, both multiple child victims under 15 (Cramer’s  $V = .08$ ) and child pornography offences (Cramer’s  $V = .14$ ) were positively significantly

associated with primary hebephilia; however, the effect sizes for both items were negligible. Despite differences in significance, these effect sizes are somewhat consistent with previous literature. For example, one study by Stephens, Seto, et al. (2017) found a negligible but significant positive association between hebephilia and the total number of victims under 15.

In terms of child pornography offences, research has indicated that there is a strong association between child pornography offending and pedophilia (Blanchard et al., 2007; Seto et al., 2006). It is important to note that Blanchard et al. (2007) and Seto et al. (2006) categorized their samples of men with pedophilia using a pedophilic index, which included individuals who had the greatest penile response to prepubescent children relative to adults but did not necessarily exclude individuals who also had a significant phallometric response to pubescent children. Therefore, there may be individuals included in their samples who are considered pedophilic, who also have hebephilic interests, given the considerable overlap between these two constructs (e.g., Stephens et al., 2017, 2019). Regardless, there has been less research that has examined child pornography offences specifically related to individuals with hebephilia; however, some research has indicated that self-reported hebephilia in child pornography offenders is associated with sexual recidivism (Seto & Eke, 2015). Non-contact sexual recidivism, which is inclusive of child pornography offences, has also been found to be associated with hebephilia (Stephens, Cantor, et al., 2017). Although these two studies did not directly analyze the association between hebephilia and child pornography offences, they lend support to the idea that child pornography offences would be associated with primary hebephilia, which was supported by this study.

Results were somewhat surprising when considering what was not associated with primary hebephilia. Three of the seven victim characteristics were not significantly associated with primary hebephilia: one or more victims between 15 and 16 (Cramer's  $V = .01$ ), the absence of gender polymorphism (Cramer's  $V = .06$ ), and extrafamilial child victims under 15 (Cramer's  $V = .02$ ). As with the significant items above, the effect sizes for these three variables were also negligible. In contrast to results by Stephens et al. (2018), which found a small significant effect between hebephilia and the number of victims who were 15 to 16 years old, the results from this study did not support the first variable. However, Stephens et al. (2018) also found that hebephilia had an association with a greater number of victims who were sexually immature (14 and younger). Although the effect sizes from both this study and Stephens et al. (2018) were small, this is consistent with literature that focuses on recidivism of perpetrators of CSA which indicates that even significant effects are usually smaller in nature (Hanson & Bussière, 1998). It is also important to note that Stephens et al. (2018) operationalized hebephilia differently than the present study. Some past studies (e.g., Stephens et al., 2018) have relied on a continuous measure of hebephilia, which contrasts with the present study which used a stricter cut-off score to categorize individuals as having primary hebephilia or not. Therefore, this more stringent definition of primary hebephilia may have affected the analyses and the association between primary hebephilia and the proposed victim characteristics.

Finally, the results did not find a significant association between the absence of gender polymorphism and primary hebephilia. The effect size for gender polymorphism was also negligible (Cramer's  $V = .06$ ), indicating that there was likely not much of a

relationship between the two variables. Instead, there was a strong association between primary hebephilia and boy victims under 15, which remained significant even when the other characteristics were accounted for. These results imply that individuals with primary hebephilia are likely to offend more against boy victims, similar to what is seen with pedophilia (Levenson et al., 2008; Seto, 2008; Seto & Lalumière, 2001). It is also worth noting that gender polymorphism was also defined differently than in other studies, which was previously mentioned in the “measures” section.

### **Creation of a Measure of Primary Hebephilia: Implications**

When accounting for the other victim characteristics, only boy victims under 15 and offences related to child pornography remained significantly associated with primary hebephilia. Given the hypotheses in the present study, these two items were combined into a scale to test its classification accuracy and examine the psychometric properties of the measure for illustrative purposes. Although the classification accuracy for these two items was weak, both items were still able to discriminate between individuals with primary hebephilia and those with no sexual interest in children, significantly just above chance level in the development group but with no significance in the validation sample.

To further examine the psychometric properties, two cut-off scores were created. When examining the psychometric properties for a cut off score of 1, primary hebephilia was detected with a sensitivity and specificity of 32% and 76%, respectively. The PPV and NPV were 41% and 68%, respectively. Using the second cut off score primary hebephilia was detected with a sensitivity and specificity of 6% and 99%, respectively. The PPV and NPV values were 69% and 67%, respectively. Therefore, the psychometric properties suggest that the SSHI pseudo measure does not provide an adequate

assessment of primary hebephilia. However, a notable limitation with the pseudo measure is that it only contains two items. Ultimately, the results from this study suggest that while no behavioural measure for assessing for primary hebephilia is available, clinicians should continue to use the SSPI-2 to assess for pedohebephilia when a sexual interest in children is possible.

It is important to also discuss possibilities as to why the proposed measure was not able to be created. The first possibility is that a behavioural measure of primary hebephilia is not possible to create. Research has demonstrated that pedophilia and hebephilia are likely more similar constructs than they are different (e.g., Stephens et al., 2019). For this reason, it may be that developing a behavioural measure for individuals with primary hebephilia is not possible; however, other measures (e.g., phallometric testing, viewing time) may be able to distinguish those with primary hebephilia from those with other types of sexual interest in children. This is an area for future research.

The second possibility is that a measure of primary hebephilia could be created using other characteristics that were not in the current dataset. For example, one item that might be associated with primary hebephilia is sexual sadism or physical violence within the context of a sexual assault. In a meta-analysis, McPhail et al. (2019) examined different kinds of stimuli used for phallometric testing to assess for sexual interests in children. Stimuli which depicted sexual violence was better at detecting hebephilia than stimuli that was passive in nature. This may be worthwhile to examine further as it could indicate, for example, that individuals who have hebephilia seek out child pornography that is violent, or that offenders who have hebephilia may have committed acts of sadism or physical violence during previous sexual assaults.

Another variable that could be used to differentiate those with primary hebephilia from those with pedophilia is charges or convictions for luring a child. In contrast to charges associated with child pornography which may arguably be most relevant to individuals who want to view prepubescent material, internet luring may be associated with hebephilia as children on the internet are typically older and have less parental supervision than younger children (Liau, Khoo, & Ang, 2008). For example, it is possible that pubescent or postpubescent teens who are using social media online are not closely monitored by caregivers, which could increase their risk of being lured. It is possible that those offenders who have hebephilic interests may create fake profiles and lure pubescent children with the intent of grooming or coercing them into sending explicit images. Therefore, it would be worthwhile to examine whether charges and convictions for internet luring are associated with primary hebephilia.

Lastly, it is also possible that changing the operationalization of primary hebephilia could have led to different findings. Originally, the author of the thesis wanted to examine exclusive hebephilia as the target group, which would have included those who met the phallometric hebephilic index cut-off only, without meeting the cut-off for the pedophilia index; however, there are challenges in defining exclusivity and no established method to do so. Future research should examine a means of operationalizing exclusive hebephilia both for research and diagnostic purposes. Although hebephilia is not listed in the *DSM-5*, it can be diagnosed under Other Specified Paraphilic Disorder (American Psychiatric Association, 2013). Clinicians can also identify whether the diagnosis is exclusive or not under a specifier, but there is no guidance or reliable

directions in place to assist clinicians in making this decision. As such, more research is needed to operationalize exclusive hebephilia and to examine its correlates.

### **Limitations**

There are several limitations of this study worth mentioning. First, it is acknowledged that the victim characteristics coded from the dataset do not accurately capture all instances of sexual offending against children, as sexual abuse is significantly underreported (Conroy & Cotter, 2017). Therefore, it is assumed this data reflects under reporting of sexual offences against children, which could have affected the SSPI-2 scores. Although nothing can be done to completely mitigate this limitation, when official and self-report sources had conflicting information, the source that had the highest number of victims was used.

A further limitation of the study includes some of the challenges that emerge when using archival data. Using an archival database did not allow the primary researchers to operationalize variables of interest since the data were already coded. For example, the variable for child pornography was coded as simply the presence or absence of a charge for making, distributing, possessing, or accessing child pornography as defined by the *Criminal Code of Canada* or if the individual admitted to accessing child pornography during the clinical interview. This variable did not consider the specific ages of the children depicted in the material, which is potentially problematic as child pornography is inclusive of teenagers who are post-pubescent. Although this may be more difficult information to glean as it is not something that is typically noted in police reports, it would be worthwhile to examine the ages of the child victims to gain a better understanding of the kind of child pornography offenders are viewing. Even better, but

less feasible, would be to have an individual who works within police services (in an ‘internet child exploitation unit,’ for example) examine the explicit material to determine what Tanner Stages the children are in, as age is an imperfect proxy for determining sexual interests in children.

There were also some limitations regarding the use of phallometric testing as the only measure to categorize offenders’ sexual interests (see chapter 1 for an exhaustive list of limitations). It is worth noting that penile responses can be suppressed, especially when the consequences of the assessment may negatively impact an individual’s life in terms of child access, sentencing, or other treatment decisions (e.g., Marshall & Fernandez, 2002). Furthermore, other measures of sexual interest were not used for this study. Viewing time (Schmidt et al., 2017) has been found to be correlated to other measures of sexual interest like self-report and phallometric testing; therefore, it would have been useful to have a second measure for assessing sexual interests. Furthermore, inter-rater reliability was not examined for any of the victim characteristics, including the SSPI-2 and additional items.

A final limitation to note is related to the phallometric stimuli used to assess participant’s sexual interests. The sets of phallometric stimuli depicted images of prepubescent, pubescent, and adults but did not include images of postpubescent teenagers approximately between the ages of 15 and 18 (see Blanchard et al., 2009 for a more thorough discussion of this limitation). Because of this limitation it is possible that there were individuals categorized in this study as having primary hebephilia who would have had substantial sexual responses to postpubescent teenagers, however there was no way to remove these individuals from the primary hebephilic group.

### **Bridging Section**

The purpose of Study 1 was to examine victim characteristics that were associated with primary hebephilia and to use them to develop a measure to assess primary hebephilia in individuals who sexually offend against children. Results from Study 1 found that a behavioural measure of primary hebephilia was unable to be developed using the hypothesized victim characteristics. Given that the original plan for Study 2 was to examine the predictive validity of the “SSH1” alongside the SSPI-2, Study 2 was amended to focus on the predictive validity of the SSPI-2. The shift to the SSPI-2 changed the focus of Study 2 to pedohebephilia as a sexual interest rather than primary hebephilia. Despite these changes, the introduction of Study 2 was largely unchanged from the proposal and only the last section of the introduction, the procedure, and hypotheses were modified to account for these changes.

### **Chapter 3**

#### **Study 2: Predictive recidivism of the Revised Screening Scale for Pedophilic Interests (SSPI-2)**

Sexual interest in children is one of the strongest predictors of sexual recidivism (Hanson & Bussière, 1998; Hanson & Morton- Bourgon, 2004; Mann et al., 2010) and an important risk factor to consider during assessment. Various assessment tools can be used to assess for a sexual interest in children (see chapter 1) and one method is the use of an individual's sexual behaviour history. Research has found that victim characteristics indicative of a sexual interest in children are also predictive of sexual recidivism (Seto et al., 2004). For example, having a male victim, younger victims, and extrafamilial victims are all associated with sexual recidivism (Seto et al., 2004; Stephens, Cantor, et al., 2017). The focus of this study was on the predictive validity of a behavioural measure of sexual interest in children.

#### **Predictive Validity of the SSPI and SSPI-2**

Much of the current literature surrounding the use of sexual behaviour history to assess sexual recidivism has focused on the predictive validity of the SSPI. The SSPI has been found to be predictive of sexual recidivism in several studies. Seto et al. (2004) found the SSPI to be positively correlated with both sexual and violent recidivism, and positively correlated with other validated risk assessment measures (e.g., Static-99, VRAG, SORAG). Their results also indicated that the SSPI was correlated with sexual recidivism among both first time and repeat adult offenders, suggesting a lengthy criminal history was not necessary to receive a high score on the SSPI (Seto et al., 2004). Another study by Helmus et al. (2015) found that the SSPI was significantly predictive of

sexual recidivism, which was determined by new charges or convictions for sexual offences. Furthermore, when compared to other validated risk assessment tools, the SSPI predicted just as well. Although some studies have found the SSPI to have good predictive validity, other studies found the opposite effect (e.g., Canales et al., 2009; Moulden et al., 2009).

There has been more limited research on the predictive validity of the SSPI-2, which was recently found to better assess pedohebephilia than hebephilia (Stephens et al., 2019). A study by Seto, Sandler, and Freeman (2017) examined the predictive and concurrent validity of the SSPI and the SSPI-2. Both measures were associated with sexual recidivism on their own (i.e., any arrest for a sexual offence within 5 years); however, the SSPI-2 was found to predict sexual recidivism slightly better than the SSPI, but the difference was not significant. Both measures were negatively associated with re-arrest at any point, and any rearrest within 5 years of release, and there was no significant relationship between the measures and violent rearrest. This indicates that the SSPI and SSPI-2 both seem to be indicators of true sexual recidivism rather than recidivism for other kinds of offences.

Research has also examined whether the SSPI and the SSPI-2 add incremental validity to other validated risk assessment measures. Incremental validity is used to examine whether a new assessment measure will increase the predictive ability beyond the existing measure. In risk assessment, using more measures does not necessarily make for a better risk prediction (Babchishin et al., 2014; Lehmann et al., 2013). Therefore, it is important to examine the incremental validity a new measure will have on an existing measure to determine whether it will be useful for clinicians to include in a battery of

assessments. In one study, Helmus et al. (2015) found that the SSPI did not add incremental validity to the Static-99R. In line with this, Seto, Sandler, and Freeman (2017) also found that neither the SSPI nor the SSPI-2 added to the Static-99R in the prediction of recidivism.

### **Hebephilia and Sexual Recidivism**

There is limited research specific to the predictive validity of hebephilia. One prominent issue is that research has used the term pedophilia to encompass different expressions of sexual interest in children, without distinguishing between individuals who are interested in pubescent or prepubescent children. Two studies have found that phallometrically assessed hebephilia was associated with sexual recidivism (e.g., Canales et al., 2009 and Proulx et al., 1997); however, it is important to note that the cut off ages for hebephilic interests in both studies were slightly older (i.e., 12 to 15 and 13 to 16, respectively). This brings into question whether the results were a true assessment of hebephilic interests, an assessment of ephebophilia (i.e., sexual interest in those who are in Tanner Stage 4; sexually maturing adolescents), or potentially a combination of both.

In a recent meta-analysis, McPhail et al. (2019) compared the predictive validity of phallometric testing for pedophilia ( $d = 0.42$ ), hebephilia ( $d = 0.43$ ), and pedohebephilia ( $d = 0.44$ ). In terms of the predictive accuracy for sexual recidivism, all three groups of offenders had relatively equivalent, small effect sizes. It is important to note that the authors used a higher cut-off age for hebephilia (ages 13 to 15) which may have affected the results. Finally, a study by Stephens, Cantor, et al. (2017) found that hebephilia was predictive of noncontact sexual recidivism (e.g., child pornography offences, exhibitionism) but that some indicators of hebephilia did not consistently

predict contact sexual reoffending. Furthermore, they found that the combination of some of the indicators (e.g., phallometric testing + self-report) better predicted sexual recidivism. Unlike the two aforementioned studies (i.e., Canales et al., 2009 and Proulx et al., 1997), phallometric testing in Stephens, Cantor, et al. (2017) accurately assessed the true nature of hebephilia, as the stimuli used were of children in Tanner Stage 2 or 3.

It could be argued that comparing a measure of primary hebephilia to a measure of pedohebephilia would produce different results. More specifically, a measure of pedohebephilia would result in a greater association with sexual recidivism because, by virtue of an individual's interests in both pubescent and prepubescent children, their potential victim pool is wider than those with only hebephilic or pedophilic interests. Therefore, in comparison to the "SSHI," which was meant to be developed to assesses primary hebephilia, the SSPI-2 which was designed to assess pedohebephilia should theoretically predict sexual recidivism better.

Although the original plan for Study 2 was to examine the predictive validity of the "SSHI," Study 1 demonstrated that a measure was not able to be developed with the proposed victim characteristics. Therefore, the focus of Study 2 shifted to examine the predictive validity of the SSPI-2, which is a measure of pedohebephilia. Study 2 has important implications because during assessments with perpetrators of CSA, one of the central factors is the risk level someone poses for sexual recidivism. For this reason, it is useful to investigate if assessment measures for sexual interest in children are indicative of true sexual recidivism. The SSPI-2 is particularly important to validate because it continues to be used by clinicians to assess for a broad sexual interest in children, which

has been found to be associated with sexual recidivism (Hanson & Morton-Bourgon, 2005).

### **Present Study**

The goal of Study 2 was to replicate findings by Seto, Sandler, and Freeman (2017) by examining whether the SSPI-2 could predict sexual recidivism in a sample of 626 perpetrators of CSA who had at least one child victim. It was hypothesized that offenders with higher scores on the SSPI-2 would be more likely to sexually recidivate than those who scored lower on the measure. Second, it was hypothesized that the SSPI-2 would not be associated with nonsexually violent and non-violent offences. Finally, the current study also examined whether the SSPI-2 would contribute to the prediction of sexual recidivism when the Static-99R was accounted for. It was hypothesized that the SSPI-2 would *not* contribute beyond the Static-99 in predicting recidivism.

### **Method**

#### **Sample**

Like Study 1, the current sample consisted of men assessed in the SBC at CAMH. The sample was initially comprised of the same 3,343 men who were referred for sexological assessments. Recidivism data was gathered from criminal records which were obtained from the Royal Canadian Mounted Police (RCMP). Criminal records were ordered for 1,117 men based on the following inclusion criteria: complete assessment information, valid phallometric testing, assessed after 1995 but before 2006 to ensure adequate follow-up time, at least one victim of any age, and the offender was 18 or older at the time of the assessment. From the 1,117 criminal records ordered, only 844 were obtained, as 273 of the records were not accessible from the national database.

Of the 844 criminal records received, only 626 were included in analyses for the current study based on further inclusion criteria: offenders must have had one child victim under the age of 15, had total SSPI-2 and Static-99R scores, and had the opportunity to reoffend in the community after the date of the initial assessment at CAMH (i.e., that they were not incarcerated for the full follow-up time). Full descriptive statistics for Study 2 variables are presented in Table 5. As with Study 1, the research questions, hypotheses, and planned analyses were also pre-registered on the Open Science Framework ([osf.io/fu3vd](https://osf.io/fu3vd)) and are outlined in Appendix B.

### **Measures**

**Recidivism.** Data for recidivism was collected through the Canadian Police Information Centre (CPIC) of the RCMP, a Canadian federal policing agency. CPIC documents included a federal identifying number for each individual, criminal charges and convictions, and the dispositions for crimes that occurred anywhere in Canada. Recidivism was coded as present for charges or convictions for new offences that occurred after the initial CAMH assessment. The decision to use charges and convictions was to capture the most accurate rate of sexual offending, as charges for sexual offences are often dropped or an offender will be convicted of a different or lesser offence that does not capture the sexual component of the incident (e.g., assault, assault with a weapon). Sexual assaults are also less likely to be reported to police than other kinds of violence (Conroy & Cotter, 2017); therefore, true rates of sexual violence are likely higher than that which is reported.

Recidivism in the original database was coded into four different categories: contact sexual offences (e.g., sexual assault, aggravated sexual assault), non-contact

sexual offences (e.g., voyeurism, exhibitionism), non-sexual violent offences (e.g., assault) and non-violent offences (e.g., condition breaches). For this study, contact and non-contact sexual offences were combined and examined as total sexual recidivism because there were low base rates of non-contact sexual offences. Recidivism included new charges or convictions against victims of any age and were dichotomously coded as being present or not for the purposes of the analyses. The dataset also included the exact end date of the follow-up period for each offender. The follow-up period ended when an offender either died or reoffended. If they reoffended, the exact date of the new charge or conviction was used. This is important because Cox regression analyses use “survival time” or the time that it takes from the point of origin until the event of interest occurs. The event of interest in this case was whether an individual incurred a new charge or conviction for reoffending.

**Opportunity to Reoffend.** Opportunity to reoffend was operationally defined as the amount of time an offender resided in the community during the follow-up period, while accounting for the time spent in secure custody. To calculate opportunity to reoffend for sexual recidivism, a composite score was calculated that involved subtracting the total number of months spent in secure custody since the start of the follow-up period from each offender’s total follow-up time for the different types of recidivism. This allowed for a more specific calculation of the opportunity each offender had to reoffend as opposed to relying solely on a set amount of follow-up time. In the Cox regression model, opportunity to reoffend represented the “time to event” or “survival time” that it took for the event of interest to occur.

**SSPI-2.** The SSPI-2 is a measure of pedohebephilic interests among male offenders who have previously committed a sexual offence against a child (Seto, Stephens, et al., 2017; Stephens et al., 2019). It consists of five items: any boy victim under the age of 15, multiple child victims under the age of 15, any child victims under the age of 12, any extrafamilial child victims under the age of 15, and any charges for child pornography offences. Scores on the SSPI-2 range from 0 to 5, with higher scores indicating a greater likelihood of pedohebephilia. Items from the SSPI-2 were scored based on victim characteristics from official file documentation. Total SSPI-2 scores were used as the predictor variable in the Cox regression model and only total scores were used for these analyses.

**Static-99R.** The Static-99R is an actuarial risk assessment used to determine risk of sexual reoffending in adult male perpetrators of CSA (Hanson & Thornton, 2002; Helmus, Thornton, et al., 2012). Eight out of the ten items are scored dichotomously: ever lived with an intimate partner, conviction for non-sexual violence at the time of index, any convictions for non-sexual violence prior to the index offence, prior sentencing dates, convictions for non-contact sexual offences, any unrelated victims, any stranger victims, and any male victims. The remaining two items are scored differently. Age at release is scored from 1 to -3, the younger the individual the higher the score. The last item, prior charges and convictions for sexual offences is scored from 0 to 3, with a higher score indicating more charges and convictions. Total Static-99R scores can range from -3 (indicating low risk of reoffending) to 12 (indicating a high risk of reoffending). The Static-99R total scores were used in the analyses. Studies have found a moderate (AUC =

.69; Helmus, Hanson, et al. 2012) to high (AUC = .74; Hanson et al., 2007) predictive validity for the Static-99R.

### **Procedure & Data Analysis**

The recidivism data was an extension of the SBC archival database described in Study 1. Criminal record files for 1,117 offenders were received for Dr. Skye Stephens' (i.e., the research supervisor of the author) dissertation on September 27, 2013 for individuals who had been assessed between 1995 and 2006. A detailed coding manual was created by Drs. Skye Stephens and Michael Seto to code the recidivism data. Both Dr. Stephens and an experienced research assistant coded the recidivism data, and coding discrepancies between the two were discussed until an agreement was reached. Inter-rater reliability was examined between the coders on 10% of the cases and ICC values for each coded variable were found to be above .90, except for violent crime which was 0.75 (Stephens et al., 2018).

To examine the first two hypotheses (the predictive validity of the SSPI-2), Cox regressions were used, and Harrell's concordance index (Harrell's C) was calculated as an effect size estimate using R (R Core Team, 2020). Harrell's C can be used as an alternative to AUCs when follow-up time is different for each offender, as was the case in the current study (Helmus & Babchishin, 2017). To examine the third hypothesis, whether the SSPI-2 contributed any additional accuracy to the Static-99R in its ability to predict recidivism, a Cox regression was also used. Research by Babchishin et al. (2012) has found that large sample sizes typically produce significant incremental validity, which was important to take into consideration for this study. Therefore, to counter this limitation, rather than analyzing both the SSPI-2 and the Static-99R in separate blocks as

typically would be done to examine true incremental validity, they were both entered together as predictors in the same block of the Cox regression.

### Results

For total sexual recidivism, the rate of recidivism was 12.1% ( $n = 76$ ). The average follow-up time for sexual recidivism was 105 months. For the other two categories, non-sexually violent recidivism and non-violent recidivism, the rates of recidivism were 11.8% ( $n = 74$ ) and 28.9% ( $n = 181$ ), respectively. The average follow-up time for nonsexually violent recidivism was 103 months and non-violent recidivism was 89 months. For the total sample, the average SSPI-2 score was 2.3 and the average Static-99R score was 1.8. There was a moderately sized positive correlation between the SSPI-2 and the Static-99R,  $r(624) = .34, p < .001$ .

As hypothesized, results revealed that total SSPI-2 scores were positively associated with an increased likelihood of sexual recidivism. The hazard ratio for sexual recidivism was 1.27, indicating that a one-point increase on the SSPI-2 was associated with a 27% increase in the hazard of sexual recidivism. Harrell's C suggested that there was a 57.7% chance that of two randomly selected cases, the one with the higher SSPI-2 score would sexually recidivate before the other. Consistent with expectations, SSPI-2 scores were not significantly associated with nonsexually violent and non-violent recidivism.

Two post hoc Cox regression analyses were conducted by considering contact and non-contact sexual recidivism separately. Results indicated that total SSPI-2 scores were significantly associated with contact sexual recidivism but not with non-contact sexual recidivism. Hazard ratios and Harrell's C values are reported in Table 6.

Lastly, results also supported the final hypothesis that the SSPI-2 would not contribute to the prediction of sexual recidivism when the Static-99R was accounted for. Results of the Cox regression indicated that the model was a significant predictor of sexual recidivism,  $\chi^2(2, 626) = 18.80, p < .001$ . While the Static-99-R contributed significantly to the model ( $B = .164, p < .001$ ), the SSPI-2 did not ( $B = .126, p = .206$ ). For the Static-99R, the hazard ratio was 1.18, indicating that a one-point increase on the Static-99R was associated with a 18% increase in the hazard of sexual recidivism. Finally, Harrell's C for the overall model was 64.2%. Refer to Table 7 for Cox regression results for the SSPI-2 and Static-99R scores.

### **Discussion**

The goal of the present study was to replicate findings by Seto, Sandler, and Freeman (2017) by examining the predictive validity of the SSPI-2 with a different sample of perpetrators of CSA. Although neither the SSPI nor the SSPI-2 were developed to assess for recidivism, studies have shown that pedophilia is associated with sexual recidivism (e.g., Helmus et al., 2015; Seto et al., 2004). Results from this study add to the literature supporting the positive association between the SSPI-2, which is a measure of pedohebephilia (Stephens et al., 2019), and sexual recidivism. More broadly these findings highlight the role that sexual interest in children plays in determining whether an individual will reoffend sexually. Although the effect size between SSPI-2 scores and sexual recidivism in this study was small, past research has also found small effects between sexual interest in children (pedohebephilia) and sexual recidivism. For example, McPhail et al. (2019) found that effect sizes for the predictive accuracy of sexual recidivism were small, and Mann et al. (2010) found the mean effect size between sexual

preference for children and the prediction of sexual recidivism to be small as well. It is important to reiterate that the SSPI-2 was not developed as a risk assessment tool and should not be used as such. Although these findings suggest that sexual interest in children plays a role in sexual recidivism, it is not the only factor that should be considered when determining whether someone will reoffend. Therefore, the SSPI-2 should continue to be used by clinicians to assess for pedohebephilia or an overall sexual interest in children but should not be the only factor taken into consideration when deciding on overall risk for recidivism.

All three hypotheses in this study were supported. First, those who scored higher on the SSPI-2 were more likely to sexually reoffend than those who scored lower on the measure. Because sexual interest in children is associated with sexual recidivism (McPhail et al., 2019) it is not surprising that those who had higher SSPI-2 scores were more likely to reoffend sexually than those who had lower SSPI-2 scores. Originally, we examined total sexual recidivism which included both contact and non-contact sexual offences because of the low base rate of sexual non-contact recidivism in the sample ( $n = 25$ ). Nonetheless, post hoc analyses were conducted for both types of sexual recidivism and only contact sexual recidivism was significantly associated with SSPI-2 scores. Consistent with Stephens et al. (2017), the effect size for contact recidivism was small. Non-contact sexual recidivism was not significantly associated with SSPI-2 scores and also produced a small effect size (Harrell's  $C = .59$ ).

Regarding hypothesis two, SSPI-2 scores were not associated with nonsexually violent and non-violent recidivism. This was expected as the SSPI-2 was not developed to assess for antisociality or general criminality, which is a stronger risk factor for non-

sexual violence in perpetrators of CSA (Brouillette-Alarie et al., 2016). Results were in line with Seto, Sandler, and Freeman (2017) as they also found that SSPI-2 scores were positively associated with sexual rearrest and negatively associated with rearrest for other reasons. Taken together, findings from the first two hypotheses add to the literature supporting the role of sexual interest in children in sexual recidivism and highlight the use of the SSPI-2 to assess for sexual interests in children.

As expected, the SSPI-2 did not contribute to the Static-99R in its ability to predict sexual recidivism. This finding was in line with Seto, Sandler, and Freeman (2017) who found that both the SSPI and the SSPI-2 did not add the Static-99R. This was also consistent with Helmus et al. (2015) who similarly found no additional accuracy when the SSPI was used alongside the Static-99R. One reason that the SSPI-2 may not be contributing predictive validity to the Static-99R is due to the overlap in constructs that exists between the two measures. Helmus et al. (2015) noted that two items from the Static-99R may be indicators of atypical sexual interests including having a male victim and having any unrelated victims. They further discuss that the difference between these items on the Static-99R and the SSPI is that the SSPI items are restricted to child victims only and the Static-99R items are scored for victims of any age. Therefore, it is not unexpected that the SSPI-2 does not contribute to the prediction of sexual recidivism above and beyond the Static-99R, as the Static-99R potentially captures atypical sexual interest.

Furthermore, the Static-99R is a comprehensive measure of risk and has items that capture other areas of risk that the SSPI-2 does not. Previous studies have found at least two factors from the items on the Static-99R including sexual criminality/deviance and

general criminality/antisocial behaviour (Allen & Pflugrad, 2014; Brouillette-Alarie & Proulx, 2013). Recently, Brouillette-Alarie et al. (2016) found three factors on the Static-99R including persistence/paraphilia, youthful stranger aggression, and general criminality; however, only youthful stranger aggression and general criminality predicted nonsexual recidivism. These factors highlight that one reason the Static-99R does a good job of assessing for recidivism is because it captures other areas of risk beyond atypical sexual interest. Therefore, it is not expected that the SSPI-2, which measures only one area of risk (sexual interest in children) would capture additional variance beyond what is already captured by the Static-99R.

Related to risk assessment, research has found that the original SSPI is associated with the deviant sexual interest items on the STABLE-2007 (Hanson et al., 2007; Helmus et al., 2015) as well as the sexual deviance items on the Violence Risk Scale-Sexual Offense Version (VRS-SO; McPhail et al., 2020). Furthermore, the SSPI can also be used on the Sex Offender Risk Appraisal Guide (SORAG) as a substitute for the phallometric section to indicate an offender's sexual preference (Seto, Sandler, & Freeman, 2017). Therefore, although not a risk assessment measure, Helmus et al. (2015) suggest that the SSPI could be used as a substitute for assessing the sexual deviance items on different risk assessment measures, such as the STABLE-2007. It might be worthwhile then, to examine whether the SSPI-2 can be used as a substitute for other sexual deviance items on risk assessment measures or whether the SSPI-2 is a stronger substitute than the SSPI for the sexual deviance items on the STABLE-2007 and the VRS-SO. Although the SSPI-2 should not be used to assess the risk of perpetrators of CSA, the results from this

study highlight the importance of sexual interest in children and sexual recidivism and the need to further examine the relationship between the two.

### **Limitations**

There are a few limitations mentioned in Study 1 that are also applicable to this study. First, that victim characteristics coded from the dataset do not capture all victims of sexual assault and therefore, could have affected overall SSPI-2 scores. Second, the SSPI-2 uses behavioural items to classify offenders as having a sexual interest in children. It would have been useful to use a second measure to also classify offenders as having a sexual interest in children when examining whether they would sexually reoffend or not. Another limitation to note is that CPIC documents were used to identify recidivism offences. Although CPIC records are meant to include all charges and convictions received anywhere in Canada, there can be discrepancies between CPIC and provincial documents. If available at the time of coding, it would have been best to use provincial documents to corroborate information from the CPIC documents. Furthermore, because CPIC was the only measure of recidivism, there could have been offences not captured officially by police and some offences were likely to have gone unreported. Therefore, true recidivism rates are likely higher than the ones coded in the dataset.

Lastly, a more diverse sample would allow for better generalizability. It would have been beneficial to also include a routine sample of perpetrators of CSA from local police data alongside the CAMH sample from this study. Another point to note is that the sample from this study was Canadian, whereas sample from Seto, Sandler, and Freeman (2017) was American. It might be beneficial to combine samples to create a larger sample more representative of North America. Finally, the sample in Seto, Sandler, and Freeman

(2017) also had higher Static-99R scores (3.52) than the offenders in this sample (1.8). A more diverse sample might also include offenders who are more antisocial, have more general violence in their offending history and therefore, potentially have higher risk assessment scores.

### **Chapter 4- General Discussion**

The assessment of sexual interest in children is important as it is a major risk factor to consider when judging the likelihood of reoffending (Hanson & Morton-Bourgon, 2005; Mann et al., 2010). More importantly, it is useful to understand whether an individual has pedophilic, hebephilic, or pedohebephilic interests to inform treatment and risk management decisions (e.g., restrictions around being in the presence of certain age groups). Therefore, the purpose of this thesis was to develop a behavioural measure to assess for primary hebephilia in perpetrators of CSA who have previously offended and to test the association of the measure with recidivism. Results from Study 1 indicated that a measure of primary hebephilia was unable to be developed with the proposed victim characteristics. Although a measure was not created, some victim characteristics (items from the SSPI-2) were positively associated with primary hebephilia. Results from this study highlight the difficulties in assessing primary hebephilia and indicate that more research is needed so clinicians can assess and classify individuals with hebephilia.

The revised purpose of Study 2 was to examine the predictive validity of the SSPI-2 with sexual, non-sexual, and non-violent recidivism. Results from this study found that the SSPI-2 was associated with sexual recidivism, but not non-sexual and non-violent recidivism. Furthermore, the SSPI-2 was no longer associated with sexual recidivism when the Static-99R was accounted for. Although the SSPI-2 was not developed to be used as a risk assessment measure and thus should not be used to assess risk, the results highlight the relationship between sexual interest in children and recidivism. Because not everyone with a sexual interest in children will offend or reoffend it is difficult to quantify how much weight should be placed on sexual

preferences when evaluating an individual's risk of reoffending. Therefore, it is important to understand other moderators that may interact with having a sexual interest in children that could increase or decrease risk for sexual recidivism.

### **Implications**

Results from this thesis have two broad implications. The first implication is that because a measure of primary hebephilia was not developed, clinicians should continue to use the SSPI-2 with the understanding that it is a measure of pedohebephilia (Stephens et al., 2019). The assessment of sexual interest in children is important for many reasons, including clinical diagnosis (Stephens & Seto, 2016). As there is no behavioural measure of hebephilia, clinicians should turn to other methods of assessing sexual interest in children. This may include phallometric testing, cognitive measures (e.g., implicit association tests and viewing time), and behavioural measures, such as the SSPI-2. The SSPI-2 should not be used as the only measure but rather in suite with other methods that assess for sexual interests. In addition to clinical diagnosis, assessing for pedohebephilia using the SSPI-2 is also related to sentencing and decisions on mandated court conditions. Offenders who score high on the SSPI-2 and who have pedohebephilia may have a broader range of potential victims. As a result, it may be beneficial for these offenders to receive specific conditions that prohibit them from attending elementary, junior high, and high schools, as well as sports and recreation facilities where both prepubescent and pubescent children frequent. Higher SSPI-2 scores might also indicate to clinicians the need for further intensive assessment (i.e., phallometric testing) to determine an offender's specific sexual interests.

Because a measure of primary hebephilia was not able to be developed it is also worth mentioning unique challenges that arise when assessing for hebephilia. First, unlike pedophilia, there are no clear diagnostic criteria for hebephilia, which can leave clinicians to their own accord when assessing individuals who they believe have hebephilic interests. As mentioned in the first chapter, the concept of hebephilia is heavily debated and there is controversy over whether it should be included in the *DSM-5* (Green, 2010; Tromovitch, 2009). Regardless of the legitimacy of hebephilia as a mental disorder, it is still a genuine sexual preference and clinicians need guidance on how to approach the assessment of hebephilia, particularly among those who have committed sexual offences.

A second implication of this thesis relates to understanding the relationship between pedohebephilia and sexual recidivism. Although sexual interest in children does not equate to sexual offending or recidivism (Kingston et al., 2007) it is an important risk factor to consider when determining an individual's risk level for recidivism among those who have previously offended. Higher scores on the SSPI-2 suggest that a sexual interest in children is likely present. Therefore, clinicians can use this information to acknowledge that this risk factor has a small positive association with recidivism (e.g., Mann et al., 2010) when reporting on the effect this may have on an offender's potential future offending. This is also important for treatment implications and behaviour management. It is crucial to directly target sexual interest in children in treatment, as it is a dynamic risk factor (Mann et al., 2010) that may be amenable with intervention.

It is also important to be cautious when interpreting the implications of the association between pedohebephilia and sexual recidivism. It is easy to presume that having a sexual interest in children is indicative of sexual offending against children, but

this is not always the case. Not all individuals who have a sexual interest in children will sexually offend against them, and not all individuals who offend against children have a sexual interest in them (Bailey et al., 2016; Kingston et al., 2007). Although it is obvious that clinicians who are tasked with assessing an individual's risk of reoffending are dealing with people who have committed at least one previous criminal offence, there is a subset of individuals who have a sexual interest in children that do not offend against children (Kingston et al., 2007; Seto, 2008). Therefore, it is worth reiterating that although having a sexual interest in children is a risk factor for future offending, it is not interchangeable with sexual offending against children. Clinicians should not assume that a sexual interest in children is a definite marker of sexual offending or reoffending for every client they work with. For this reason, it would also be worthwhile to develop a measure for assessing sexual interest in children that does not rely on victim characteristics, especially for individuals who have never sexually offended before but may be at risk of offending.

### **Future Directions**

Although the present thesis was not able to develop a measure of primary hebephilia, future research should not be discouraged in trying to create both a behavioural measure and other measures to assess for primary hebephilia. For example, cognitive measures have been used with perpetrators of CSA to determine whether a sexual interest in children is present (Banse et al., 2010; Gray et al., 2005). Specific cognitive measures like Implicit Association Tasks (IAT; Brown et al., 2009) have found significant group differences between pedophilia, hebephilia, and a control group in associating children and sex. Therefore, it is worthwhile that future research continues to examine the

association between IATs and hebephilia and whether IATs may be able to accurately identify those with primarily hebephilic interest. Another area for future research would be to examine whether those with pedohebephilia have higher recidivism than those with primary pedophilia or primary hebephilia.

Researchers should also consider the importance of standardizing terms related to sexual interest in children. As discussed in chapter 1, 'exclusivity' of sexual interests is a difficult concept to assess as there is no objective definition or instructions for clinicians on how and when to classify someone as having an exclusive sexual interest. Although there is no concrete definition of exclusivity, research has operationalized this in various ways and found that people with hebephilia have higher rates of non-exclusivity (Beier et al., 2015). This is important as it has implications for treatment goals of individuals with sexual interest in children. For example, if an individual has non-exclusive hebephilia, the sexual response gradient posits that they may also have interests in the age category above hebephilia, to young adults (Seto et al., 1999). Therefore, individuals with non-exclusive hebephilia may have a more positive prognosis since they would have some interest in adults which could be explored further in treatment.

It is also important that future research operationalize exclusivity for both research and assessment/diagnostic purposes. Along with objective definitions, it is also important to note that there are inconsistencies in grouping perpetrators of CSA into specific erotic age preference categories. This has implications for research when developing tools or measures to assess for a sexual interest in children. Some studies group perpetrators of CSA into having pedophilia, hebephilia, or pedohebephilia using their highest phallometric responses (e.g., Blanchard et al., 2007 and Seto et al., 2006), while others

rely on a continuous measure of categorization (e.g., Stephens et al., 2018). When using phallometric testing to categorize offenders, crossover or overlap between age erotic categories typically exists, as demonstrated by the data used for this thesis. Creating an objective definition and means of categorizing offenders based on their age preferences would be beneficial as it would allow for generalizability and consistency among research studies.

### References

- Allen, M. & McCarthy, K. (2018). *Victims of police-reported violent crime in Canada: National, provincial, and territorial fact sheets, 2016*. Catalogue no. 85-002-X, ISSN 1209-6393. <https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2018001/article/54960-eng.pdf?st=8caU1wPS>
- Allen, B. P., & Pflugrad, D. M. (2014). An exploration of the latent constructs of the Static-99. *International Journal of Offender Therapy and Comparative Criminology*, 58(11), 1376-1388. <https://doi.org/10.1177/0306624X13496046>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. <https://doi.org/10.1176/appi.books.9780890425596>
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment*, 19(4), 442-461. <https://doi.org/10.1177/1073191112458312>
- Babchishin, K. M., Nunes, K. L., & Kessous, N. (2014). A multimodal examination of sexual interest in children: A comparison of sex offenders and nonsex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 26(4), 343-374. <https://doi.org/10.1177/1079063213492343>
- Bailey, J. M., Bernhard, P. A., & Hsu, K. J. (2016). An internet study of men sexually attracted to children: Correlates of sexual offending against children. *Journal of Abnormal Psychology*, 125(7), 989-1000. <http://dx.doi.org/10.1037/abn0000213>

- Banase, 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), 319-355. <https://doi.org/10.1177/0093854809357598>
- Barth, J., Bermetz, L., Heim, E., Trelle, S., & Tonia, T. (2013). The current prevalence of child sexual abuse worldwide: A systematic review and meta-analysis. *International Journal of Public Health*, 58(3), 469-483. <https://doi.org/10.1007/s00038-012-0426-1>
- Baxter, D. J., Marshall, W. L., Barbaree, H. E., Davidson, P. R., & Malcom, P. B. (1984). Deviant sexual behaviour: Differentiating sex offenders by criminal and personal history, psychometric measures, and sexual response. *Criminal Justice and Behaviour*, 11(4), 477- 501. <https://doi.org/10.1177/0093854884011004007>
- Beier, K. M., Amelung, T., Kuhle, L., Grundmann, D., Scherner, G., & Neutze, J. (2013). Hebephilia as a sexual disorder. *Fortschritte der Neurologie-Psychiatrie*, 81(3), 128-137. <https://doi.org/10.1055/s-0032-1330539>
- Beier, K. M., Grundmann, D., Kuhle, L. F., Scherner, G., Konrad, A., & Amelung, T. (2015). The German Dunkelfeld project: A pilot study to prevent child sexual abuse and the use of child abusive images. *Journal of Sexual Medicine*, 12(2), 529-542. <https://doi.org/10.1111/jsm.12785>
- Beier, K. M., Neutze, J., Mundt, I. A., Ahlers, C. J., Goecker, D., Konrad, A., Schaefer, G. A. (2009). Encouraging self-identified pedophiles and hebephiles to seek professional help: First results of the Prevention Project Dunkelfeld (PPD). *Child Abuse & Neglect*, 33(8), 545-549. <https://doi.org/10.1016/j.chiabu.2009.04.002>

Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders.

*Psychological Assessment, 13*(1), 118-126. <https://doi.org/10.1037/1040-3590.13.1.118>

Blanchard, R., Kolla, N. J., Cantor, J. M., Klassen, P. E., Dickey, R., Kuban, M. E., & Blak, T. (2007). IQ, handedness, and pedophilia in adult male patients stratified by referral source. *Sexual Abuse, 19*(3), 285-309.

<https://doi.org/10.1177/107906320701900307>

Blanchard, R., Kuban, M. E., Blak, T., Klassen, P. E., Dickey, R., & Cantor J, M. (2012). Sexual attraction to others: A comparison of two models of alloerotic responding in men. *Archives of Sexual Behavior, 41*(1), 13-29.

<https://doi.org/10.1007/s10508-010-9675-3>

Blanchard, R., Lykins, A. D., Wherrett, D., Kuban, M. E., Cantor, J. M., Blak, T.,

Dickey, R., & Klassen, P. E. (2009). Pedophilia, hebephilia, and the DSM-V.

*Archives of Sexual Behavior, 38*(3), 335-350. <https://doi.org/10.1007/s10508-008-9399-9>

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender*

*assessment and rehabilitation* (User Report 2007–06). Ottawa, Public Safety

Canada. <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/rsk-nd-rspnsvty/rsk-nd-rspnsvty-eng.pdf>

Brouillette-Alarie, S., Babchishin, K. M., Hanson, K. R., Helmus, L. M. (2016). Latent constructs of the Static-99R and Static-2002R: A three-factor solution.

*Assessment, 23*(1), 96-111. <https://doi.org/10.1177/1073191114568114>

- Brouillette-Alarie, S., & Proulx, J. (2013). Predictive validity of the Static-99R and its dimensions. *Journal of Sexual Aggression, 19*(3), 311-328.  
<https://doi.org/10.1080/13552600.2012.747630>
- Brown, A. S., Gray, N. S., & Snowden, R. J. (2009). Implicit measurement of sexual associations in child sex abusers: Role of victim type and denial. *Sexual Abuse, 21*(2), 166–180. <https://doi.org/10.1177/1079063209332234>
- Canales, D. D., Olver, M. E., & Wong, S. C. P. (2009). Construct validity of the violence risk scale-sexual offender version for measuring sexual deviance. *Sexual Abuse, 21*(4), 474-492. <https://doi.org/10.1177/1079063209344990>
- Cantor, J. M., Blanchard, R., Christensen, B. K., Dickey, R., Klassen, P. E., Beckstead, A. L., Blak, T., & Kuban, M. E. (2004). Intelligence, memory, and handedness in pedophilia. *Neuropsychology, 18*(1), 3-14. <https://doi.org/10.1037/0894-4105.18.1.3>
- Cantor, J. M., & McPhail, I. V. (2015). Sensitivity and specificity of the phallometric test for hebephilia. *The Journal of Sexual Medicine, 12*(9), 1940-1950.  
<https://doi.org/10.1111/jsm.12970>
- Conroy, S., & Cotter, A. (2017). Self-reported sexual assault in Canada, 2014. Catalogue no. 85-002-X, ISSN 1209- 6393. <https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2017001/article/14842-eng.pdf?st=b2FFW9jw>
- Chivers, M. L., Seto, M. C., Lalumière, M. L., Laan, E., & Grimbos, T. (2010). Agreement of self-reported and genital measures of sexual arousal in men and women: A meta-analysis. *Archives of Sexual Behavior, 39*(1), 5-56.  
<https://doi.org/10.1007/s10508-009-9556-9>

Coric, V., Feuerstein, S., Fortunati, F., Southwick, S., Temporini, H., & Morgan, C. A.

(2005). Assessing sex offenders. *Psychiatry (Edgmont)*, 2(11), 26-29.

[https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993520/pdf/PE\\_2\\_11\\_26.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2993520/pdf/PE_2_11_26.pdf)

Feldman, P., & MacCulloch, M. (1980). *Human sexual behaviour*. John Wiley.

Finkelhor, D. & Shattuck, A. (2012). *Characteristics of crimes against juveniles*. Crimes against Children Research Center.

[http://www.unh.edu/ccrc/pdf/CV26\\_Revised%20Characteristics%20of%20Crimes%20against%20Juveniles\\_5-2-12.pdf](http://www.unh.edu/ccrc/pdf/CV26_Revised%20Characteristics%20of%20Crimes%20against%20Juveniles_5-2-12.pdf)

Franklin, K. (2009). The public policy implications of “hebephilia”: A response to

Blanchard et al. 2008. *Archives of Sexual Behaviour*, 38(3), 319–320.

<https://doi.org/10.1007/s10508-008-9425-y>

Franklin, K. (2010). Hebephilia: Quintessence of diagnostic pretextuality. *Behavioral*

*Sciences & the Law*, 28(6), 751-768. <https://doi.org/10.1002/bsl.934>

Frenzel, R. R., & Lang, R. A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child sexual abusers. *Annals of Sex Research*, 2(3), 255-275.

<https://doi.org/10.1007/BF00849718>

Freund, K. (1957). Diagnostika homosexuality u muzu [diagnosing homosexuality in men]. *Ceskoslovak Medicine*, 53, 382-393.

Freund, K. (1965). Diagnosing heterosexual pedophilia by means of a test for sexual interest. *Behaviour Research and Therapy*, 3(4), 229-234.

[https://doi.org/10.1016/0005-7967\(65\)90031-8](https://doi.org/10.1016/0005-7967(65)90031-8)

Freund, K., & Blanchard, R. (1989). Phallometric diagnosis of pedophilia. *Journal of Consulting and Clinical Psychology, 57*(1), 100-105.

<https://doi.org/10.1037/0022-006X.57.1.100>

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. <https://doi.org/10.1037/1040-3590.3.2.254>

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. <https://doi.org/10.1037/0021-843X.114.2.304>

Green, R. (2010). Sexual preference for 14-year-olds as a mental disorder: You can't be serious!! *Archives of Sexual Behaviour, 39*(3), 585–586.

<https://doi.org/10.1007/s10508-010-9602-7>

Guay, J., Proulx, J., Cusson, M., & Ouimet, M. (2001). Victim-choice polymorpha among serious sex offenders. *Archives of Sexual Behavior, 30*(5), 521-533.

<https://doi.org/10.1023/A:1010291201588>

Hames, R., & Blanchard, R. (2012). Anthropological data regarding the adaptiveness of hebephilia. *Archives of Sexual Behaviour, 41*(4), 745-747.

<https://doi.org/10.1007/s10508-012-9972-0>

Hammer, E. F., & Glueck, B. C. (1955). Psychodynamic patterns in the sex offender: Fear of the adult female sex object and feelings of genital inadequacy. In P. H.

Hoch and J. Zubin (Eds.), *Psychiatry and the law* (pp. 157-168). Grune and Stratton.

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348–362. <https://doi.org/10.1037/0022-006X.66.2.348>

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 No. 2007–05). Ottawa, Public Safety Canada.  
<https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/ssssng-rsk-sxl-ffndrs-eng.pdf>

Hanson, R. K., & Morton-Bourgon, K. E. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Cat. No. PS3-1/2004-2). Ottawa, Public Safety and Emergency Preparedness Canada.  
<https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/2004-02-prdctrs-sxl-rcdvsm-pdtd/2004-02-prdctrs-sxl-rcdvsm-pdtd-eng.pdf>

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24(1), 119–136.  
<https://doi.org/10.1023/A:1005482921333>

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, 73(6), 1154-1163. <https://doi.org/10.1037/0022-006X.73.6.1154>

- Heil, P., Ahlmeyer, S., & Simons, D. (2003). Crossover sexual offenses. *Sexual Abuse, 15*(4), 221–236. <https://doi.org/10.1177/107906320301500401>
- Heiman, J. R. (1978). Uses of psychophysiology in the assessment and treatment of sexual dysfunction. In J. LoPiccolo & L. LoPiccolo (Eds.), *Handbook of Sex Therapy* (pp. 123-135). Springer. [https://doi.org/10.1007/978-1-4613-3973-1\\_8](https://doi.org/10.1007/978-1-4613-3973-1_8)
- Helmus, L. M., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*(1), 8-25. <https://doi.org/10.1177/0093854816678898>
- Helmus, L., Ó Ciardha, C., & Seto, M. C. (2015). The screening scale for pedophilic interests (SSPI): Construct, predictive, and incremental validity. *Law and Human Behavior, 39*(1), 35-43. <https://doi.org/10.1037/lhb0000099>
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse, 24*(1), 64-101. <https://doi.org/10.1177/1079063211409951>
- Kingston, D. A., Firestone, P., Moulden, H. M., & Bradford, J. M. (2007). The utility of the diagnosis of pedophilia: A comparison of various classification procedures. *Archives of Sexual Behavior, 36*(3), 423-436. <https://doi.org/10.1007/s10508-006-9091-x>
- Kleban H., Chesin, M. S., Jeglic, E. L., & Mercado, C. C. (2013). An exploration of crossover sexual offending. *Sexual Abuse, 25*(5), 427-443. <https://doi.org/10.1177/1079063212464397>

- Konrad, A., Kuhle, L. F., Amelung, T., & Beier, K. M. (2018). Is emotional congruence with children associated with sexual offending in pedophiles and hebephiles from the community? *Sexual Abuse, 30*(1), 3–22. <https://doi.org/10.1177/1079063215620397>
- Laws, D. R. (2009). Penile plethysmography: Strengths, limitations, innovations. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offenders* (pp. 7-29). John Wiley & Sons Ltd. <https://doi.org/10.1002/9780470747551>
- Lalumière, M. L., & Quinsey, V. L. (1994). The discriminability of rapists from non-sex offenders using phallometric measures: A meta-analysis. *Criminal Justice and Behavior, 21*(1), 150-175. <https://doi.org/10.1177/0093854894021001010>
- Lehmann, R. J. B., Hanson, K. R., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K. P. (2013). Interpreting multiple risk scales for sex offenders: Evidence for averaging. *Psychological Assessment, 25*(3), 1019-1024. <https://doi.org/10.1037/a0033098>
- Levenson, J. S., Becker, J., & Morin, J. W. (2008). The relationship between victim age and gender crossover among sex offenders. *Sexual Abuse, 20*(1), 43-60. <https://doi.org/10.1177/1079063208314819>
- Liau, A. K., Khoo, A., & Ang, P. H. (2008). Parental awareness and monitoring of adolescent internet use. *Current Psychology, 27*(4), 217-233. <https://doi.org/10.1007/s12144-008-9038-6>

Maltz, W. (2002). Treating the sexual intimacy concerns of sexual abuse survivors.

*Sexual and Relationship Therapy, 17*(4), 321-327.

<https://doi.org/10.1080/1468199021000017173>

Maniglio, R. (2009). The impact of child sexual abuse on health: A systematic review of reviews. *Clinical Psychology Review, 29*(7), 647-657.

<https://doi.org/10.1016/j.cpr.2009.08.003>

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism:

Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse, 22*(2), 191-217. <https://doi.org/10.1177/1079063210366039>

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.), *Applied clinical psychology. Handbook of sexual assault: Issues, theories, and treatment of the offender* (pp. 257–275). Springer. <https://doi.org/10.1007/978-1-4899-0915-2>

Marshall, W. L., & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders:

Limits to its value. *Clinical Psychology Review, 20*(7), 807-822.

[https://doi.org/10.1016/S0272-7358\(99\)00013-6](https://doi.org/10.1016/S0272-7358(99)00013-6)

McPhail, I. V., Hermann, C. A., Fernane, S., Fernandez, Y. M., Nunes, K. L., & Cantor, J.

M. (2019). Validity in phallometric testing for sexual interests in children: A meta-analytic review. *Assessment, 26*(3), 535-551.

<https://doi.org/10.1177/1073191117706139>

- McPhail, I. V., Olver, M. E., Nicholaichuk, T. P., & Haynes, A. (2020). Convergent and predictive associations of three measures of pedophilic interest. *Sexual Abuse*.  
<https://doi.org/10.1177/1079063220968042>
- 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). Wiley-Blackwell.  
<https://doi.org/10.1002/9781119990420.ch7>
- Mokros, A., Gebhard, M., Heinz, V., Marschall, R. W., Nitschke, J., Glasgow, D. V., Gress, C. L. Z., Laws, R. (2013). Computerized assessment of pedophilic sexual interest through self-report and viewing time: Reliability, validity, and classification accuracy of the affinity program. *Sexual Abuse*, 25(3), 230-258.  
<https://doi.org/10.1177/1079063212454550>
- Money, J. (1986). *Lovemaps: Clinical concepts of sexual/erotic health and pathology, paraphilia, and gender transposition in childhood, adolescence, and maturity*. Irvington Publishers.
- Moss S., Simmons M., Trendell S., Stephens S. (2018) Assessing and treating men who have committed sexual offenses. In M, Ternes, P.R. Magaletta, & M.W. Patry (Eds.), *The practice of correctional psychology* (pp. 197-218). Springer, Cham. [https://doi.org/10.1007/978-3-030-00452-1\\_9](https://doi.org/10.1007/978-3-030-00452-1_9)
- Moulden, H. M., Firestone, P., Kingston, D., & Bradford, J. (2009). Recidivism in pedophiles: An investigation using different diagnostic methods. *The Journal*

*of Forensic Psychiatry & Psychology*, 20(5), 680-701.

<https://doi.org/10.1080/14789940903174055>

Murray, L.K., Nguyen, A., & Cohen, J. A. (2014). Child sexual abuse. *Child and Adolescent Psychiatric Clinics of North America*, 23(2), 321–337.

<https://doi.org/10.1016/j.chc.2014.01.003>

Nichols, H.R., & Molinder, I. (2000). *Multiphasic Sex Inventory II*. Nichols & Molinder Assessments, Inc. <https://www.nicholsandmolinder.com/sex-offender-assessment-msi-ii-am.php>

Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J., & Ouimet, M. (1997). Static and dynamic predictors of recidivism in sexual aggressors. *Sexual Abuse*, 9(1), 7-27. <https://doi.org/10.1007/BF02674889>

R Core Team (2020). R: A language and environment for statistical computing. (Version 4.0.2) R foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Ratican, K. L. (1992). Sexual abuse survivors: Identifying symptoms and special treatment considerations. *Journal of Counseling & Development*, 71(1), 33-38. <https://doi.org/10.1002/j.1556-6676.1992.tb02167.x>

Rempel, J. K., & Serafini, T. E. (1995) Factors influencing the activities that people experience as sexually arousing: A theoretical model. *Canadian Journal of Human Sexuality*, 4(1), 3-14.

Rice, M. E., Harris, G. T., Lang. C., & Cormier, C. (2006). Violent sex offences: How are the best measured from official records? *Law and Human Behavior*, 30(4), 525-541. <https://doi.org/10.1007/s10979-006-9022-3>

- Rind, B., & Yuill, R. (2012). Hebephilia as mental disorder? A historical, cross-cultural, sociological, cross-species, non-clinical empirical and evolutionary review. *Archives of Sexual Behaviour, 41*(4), 797–829. <https://doi.org/10.1007/s10508-012-9982-y>
- Rosenzweig, S. (1942). The photoscope as an objective device for evaluating sexual interest. *Psychosomatic Medicine, 4*(2), 150–158.
- Schmidt, A. F., Babchishin, K. M., & Lehmann, R. J. B. (2017). A meta-analysis of viewing time measures of sexual interest in children. *Archives of Sexual Behavior, 46*(1), 287-300. <https://doi.org/10.1007/s10508-016-0806-3>
- Schmidt, A. F. & Perkins, D. (2020). Using the explicit and implicit sexual interest profile in applied forensic or clinical contexts. In G. Ackerman, D. Perkins & R. Bartels (Eds.), *Assessing and managing problematic sexual interests: A practitioner's guide*. Routledge. <https://doi.org/10.4324/9780429287695>
- Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. American Psychological Association. <https://doi.org/10.1037/11639-000>
- Seto, M. C. (2009). Pedophilia. *Annual Review of Clinical Psychology, 5*(1), 391–407. <https://doi.org/10.1146/annurev.clinpsy.032408.153618>
- Seto, M. C. (2013). *Internet sex offenders*. American Psychological Association. <https://doi.org/10.1037/14191-000>
- Seto, M. C. (2016). The puzzle of male chronophilias. *Archives of Sexual Behavior, 46*(1), 3-22. <https://doi.org/10.1007/s10508-016-0799-y>

- Seto, M. C. (2018). *Pedophilia and sexual offending against children: Theory, assessment, and intervention* (2<sup>nd</sup> ed.). American Psychological Association.  
<https://doi.org/10.1037/11639-000>
- Seto, M. C. (2019). The motivation-facilitation model of sexual offending. *Sexual Abuse, 31*(1) 3-24. <https://doi.org/10.1177/1079063217720919>
- Seto, M. C., Buckman, C., Dwyer, R. G., & Quayle, E. (2018). *Production and active trading of child sexual exploitation images depicting identified victims*.  
[https://www.missingkids.org/content/dam/missingkids/pdfs/nmec-analysis/Production%20and%20Active%20Trading%20of%20CSAM\\_FullReport\\_FINAL.pdf](https://www.missingkids.org/content/dam/missingkids/pdfs/nmec-analysis/Production%20and%20Active%20Trading%20of%20CSAM_FullReport_FINAL.pdf)
- Seto, M. C., Cantor, J. M., & Blanchard, R. (2006). Child pornography offenses are a valid diagnostic indicator of pedophilia. *Journal of Abnormal Psychology, 115*(3), 610–615. <https://doi.org/10.1037/0021-843X.115.3.610>
- Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The screening scale for pedophilic interests and recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior, 33*(5), 455-466.  
<https://doi.org/10.1023/b:aseb.0000037426.55935.9c>
- Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse, 13*(1), 15-25.  
<https://doi.org/10.1177/107906320101300103>
- Seto, M. C., Lalumière, M. L., & Kuban, M. (1999). The sexual preferences of incest offenders. *Journal of Abnormal Psychology, 108*(2), 267-272.  
<https://doi.org/10.1037/0021-843X.108.2.267>

- Seto, M. C., Sandler, J. C., & Freeman, N. J. (2017). The revised screening scale for pedophilic interests: Predictive and concurrent validity. *Sexual Abuse, 29*(7), 636-657. <https://doi.org/10.1177/1079063215618375>
- Seto, M. C., Stephens, S., Lalumière, M. L., & Cantor, J. M. (2017). The revised screening scale for pedophilic interests (SSPI-2): Development and criterion-related validation. *Sexual Abuse, 29*(7), 619-635. <https://doi.org/10.1177/1079063215612444>
- Singer, B. (1984). Conceptualizing sexual arousal and attraction. *The Journal of Sex Research, 20*(3), 230-240. <https://doi.org/10.1080/00224498409551222>
- Stephens, S., Cantor, J. M., Goodwill, A. M., & Seto, M. C. (2017). Multiple indicators of sexual interest in prepubescent or pubescent children as predictors of sexual recidivism. *Journal of Consulting and Clinical Psychology, 85*(6), 585-595. <https://doi.org/10.1037/ccp0000194>
- Stephens, S., & Seto, M. C. (2016). Hebephilic sexual offending. In A. Phenix, H. Hoberman (Eds.), *Sexual Offending: Predisposing antecedents, assessments, and management* (pp. 29-43). Springer. [https://doi.org/10.1007/978-1-4939-2416-5\\_3](https://doi.org/10.1007/978-1-4939-2416-5_3)
- Stephens, S., Seto, M. C., Cantor, J. M., Lalumière, M. L. (2019). The revised screening scale for pedophilic interests-revised (SSPI-2) may be a measure of pedohebephilia. *The Journal of Sexual Medicine, 16*(10), 1655-1663. <https://doi.org/10.1016/j.jsxm.2019.07.015>

- Stephens, S., Seto, M. C., Goodwill, A. M., & Cantor, J. M. (2017). Evidence of construct validity in the assessment of hebephilia. *Archives of Sexual Behaviour*, 46(1), 301-309. <https://doi.org/10.1007/s10508-016-0907-z>
- Stephens, S., Seto, M.C., Goodwill, A. M., & Cantor, J.M. (2018). Age diversity among victims of hebephilic sexual offenders. *Sexual Abuse*, 30(3), 322-339. <https://doi.org/10.1177/1079063216665837>
- Stoltenborgh, M., Van IJzendoorn, M. H., Euser, E. M., & Bakermans-Kranenburg, M. J. (2011). A global perspective on child sexual abuse: Meta-Analysis of prevalence around the world. *Child Maltreatment*, 16(2), 79-101. <https://doi.org/10.1177/1077559511403920>
- Townsend, C. (2013). *Prevalence and consequences of child sexual abuse compared with other childhood experiences*. Charleston, S.C., Darkness to Light. [www.D2L.org](http://www.D2L.org).
- Tozdan, S., & Briken, P. (2015). The earlier, the worse? Age of onset of sexual interest in children. *Journal of Sexual Medicine*, 12(7), 1602-1608. <https://doi.org/10.1111/jsm.12927>
- Tromovitch, P. (2009). Manufacturing mental disorder by pathologizing erotic age orientation: A comment on Blanchard et al. (2008). *Archives of Sexual Behaviour*, 38(3), 328. <https://doi.org/10.1007/s10508-008-9426-x>
- Westlake, B.G. (2020). The past, present, and future of online child sexual exploitation: Summarizing the evolution of production, distribution, and detection. In T.J Holt & A.M Bossler (Eds.), *The palgrave handbook of international cybercrime and cyberdeviance* (pp. 1225-1253). Palgrave Macmillan, Cham. [https://doi.org/10.1007/978-3-319-78440-3\\_52](https://doi.org/10.1007/978-3-319-78440-3_52)

Wilson, R. J., & Miner, M. H. (2016). Measurement of male sexual arousal and interest using penile plethysmography and viewing time. In D. R. Laws & W. O'Donohue (Eds.), *Treatment of sex offenders: Strengths and weaknesses in assessment and intervention* (pp. 107-131). Springer. [https://doi.org/10.1007/978-3-319-25868-3\\_5](https://doi.org/10.1007/978-3-319-25868-3_5)

Wormith, J. S. (1986). Assessing deviant sexual arousal: Psychological and cognitive aspects. *Advances in Behaviour Research and Therapy*, 8(3), 101-137.

**Table 1***Descriptive Statistics for Study 1 Variables (N = 1,655)*

Demographic variables	Primary hebephilia ( <i>n</i> = 567)	No sexual interest ( <i>n</i> = 1,088)	Total sample ( <i>n</i> = 1,655)
Average age at assessment ( <i>n</i> = 1,655)	39.8 (14.2)	38.5 (13.6)	38.9 (13.8)
Average age first known offence ( <i>n</i> = 1,383)	29.0 (13.7)	29.2 (12.9)	29.2 (13.2)
Average age at last known offence ( <i>n</i> = 1,384)	36.4 (13.7)	35.0 (13.2)	35.5 (13.4)
Married or common law ( <i>n</i> = 1,393)	332 (69.5%)	660 (72.1%)	992 (71.2%)
No known psychiatric history ( <i>n</i> = 1,115)	348 (88.1%)	652 (90.6%)	1000 (89.7%)
Ethnicity ( <i>n</i> = 1,647)			
Caucasian	477 (84.7%)	860 (79.3%)	1337 (81.2%)
African Canadian	22 (3.9%)	68 (6.3%)	90 (5.5%)
East Indian or Pakistani	16 (2.8%)	40 (3.7%)	56 (3.4%)
Aboriginal Canadian	15 (2.7%)	36 (3.3%)	51 (3.1%)
Asian Canadian	3 (0.5%)	18 (1.7%)	21 (1.3%)
Filipino or Pacific Islander	8 (1.4%)	13 (1.2%)	21 (1.3%)
Other	22 (3.9%)	49 (4.5%)	71 (4.3%)
Referral ( <i>n</i> = 1,141)			
Probation or parole	170 (42.4%)	350 (47.3%)	520 (45.6%)
Correctional institution	126 (31.4%)	222 (30.0%)	348 (30.5%)
Own lawyer	68 (17.0%)	128 (17.3%)	196 (17.2%)
Self-referral	35 (8.7%)	35 (4.7%)	70 (6.1%)
Legal Aid	2 (0.5%)	5 (0.7%)	7 (0.6%)
SSPI-2 items ( <i>n</i> = 1,655)			
1. Any boy victim under 15	180 (31.7%)	260 (23.9%)	440 (26.6%)
2. Multiple child victims under 15	323 (57.0%)	489 (44.9%)	812 (49.1%)
3. Any child victim under 12	410 (72.3%)	689 (63.3%)	1,099 (66.4%)
4. Any extrafamilial child victims under 15	463 (81.7%)	832 (76.5%)	1,295 (78.2%)
5. Any possession of child pornography	86 (15.2%)	67 (6.2%)	153 (9.2%)
Additional items ( <i>n</i> = 1,655)			
One or more victim 15/16	77 (13.6%)	130 (11.9%)	207 (12.5%)
Gender polymorphism	93 (16.4%)	125 (11.5%)	218 (13.2%)
Average pedophilia index score	0.2 (0.5)	-0.9 (0.7)	-0.5 (0.8)
Average hebephilia index score	0.8 (0.4)	-0.6 (0.6)	-0.1 (0.8)

*Note.* For continuous variables, means and standard deviations (in parentheses) are reported and for

categorical variables, frequencies and percentages (in parentheses) are reported.

**Table 2***Frequency and Chi Square Results for the Development Sample (N = 827)*

Victim items	Primary hebephilia (n = 283)	No sexual interest (n = 544)	$\chi^2$	Cramér's V
SSPI-2 items				
1. Any boy victim under 15	93 (32.9%)	127 (23.3%)	8.64**	.10
2. Multiple child victims under 15	150 (53.0%)	244 (44.9%)	4.96*	.08
3. Any child victim under 12	202 (71.4%)	346 (63.6%)	5.03*	.08
4. Any extrafamilial child victims under 15	224 (79.2%)	419 (77.0%)	0.49	.02
5. Any possession of child pornography	42 (14.8%)	35 (6.4%)	15.58***	.14
Additional items				
One or more victim 15/16	32 (11.3%)	64 (11.8%)	0.04	.01
Gender polymorphism	44 (15.5%)	61 (11.2%)	3.16	.06

Note. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

**Table 3***Logistic Regression with the Significant Predictors for the Development Sample (N = 827)*

Indicator	B (SE)	Wald	Odd's Ratio [95% CI]
Item 1: Any boy victim under 15	0.36 (0.17)	4.34*	1.43 [1.02, 2.00]
Item 2: Multiple child victims under 15	0.18 (0.16)	1.21	1.19 [0.87, 1.63]
Item 3: Any child victim under 12	0.28 (0.17)	2.82	1.33 [0.95, 1.85]
Item 5: Any possession of child pornography	0.94 (0.25)	14.62***	2.55 [1.58, 4.12]
Intercept	-1.13 (0.15)	58.84***	0.33

*Note.* The dependent variable was hebephilic interests with primary hebephilia as the target category and no sexual interest in children as the reference category.

Nagelkerke  $R^2 = .045$

SE = standard error; CI = confidence interval

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

**Table 4**

*Psychometric Properties for Different “SSHI” Cut-Off Scores for Primary Hebephilia vs. Not Having a Sexual Interest in Children (N = 1,655) (both indices  $z > 0.25$ )*

“SSHI” Score	% with Primary Hebephilia	Odds Ratio [95% CI]	SE	SP	PPV	NPV
$\geq 1$ (n = 440)	40.9%	1.48 [1.18, 1.86]	32%	76%	41%	68%
= 2 (n = 49)	69.4%	4.56 [2.46, 8.45]	6%	99%	69%	67%

*Note.* SE = sensitivity; SP = specificity; PPV = positive predictive value; NPV = negative predictive value; SSHI = Screening Scale for Hebephilic Interests; CI = confidence interval

The number people classified as having primary hebephilia with a “SSHI” score of zero was 387 (31.9%)

**Table 5***Descriptive Statistics for Study 2 Variables*

Demographic variables	<i>M(SD)</i> or <i>n(%)</i>
Average age at assessment ( <i>n</i> = 626)	41.3 (13.0)
Average age first known offence ( <i>n</i> = 621)	30.4 (13.0)
Average age at last known offence ( <i>n</i> = 622)	37.5 (12.6)
Married or common law ( <i>n</i> = 625)	475 (76.0%)
No known psychiatric history ( <i>n</i> = 596)	542 (90.9%)
Ethnicity ( <i>n</i> = 622)	
Caucasian	508 (81.7%)
African Canadian	31 (5.0%)
East Indian or Pakistani	14 (2.3%)
Aboriginal Canadian	20 (3.2%)
Asian Canadian	7 (1.1%)
Filipino or Pacific Islander	12 (1.9%)
Other Ethnicity	30 (4.8%)
Referral ( <i>n</i> = 609)	
Probation or parole	285 (46.8%)
Correctional institution	191 (31.4%)
Own lawyer	118 (19.4%)
Self- referral	9 (1.5%)
Legal Aid	6 (1.0%)

*Note.* For continuous variables, means and standard deviations (in parentheses) are

reported and for categorical variables, frequencies and percentages (in parentheses) are reported.

**Table 6***SSPI-2 Cox Regression Results for Predicting Different Recidivism (N = 626)*

Predictor	B (SE)	Wald	HR [95% CI]	Harrell's C
Total sexual recidivism	0.24 (0.09)	6.53**	1.27 [1.06, 1.52]	.577
Contact sexual recidivism	0.23 (0.10)	4.92**	1.26 [1.03, 1.53]	.578
Non-contact sexual recidivism	0.31 (0.17)	3.55	1.36 [0.99, 1.89]	.591
Nonsexually violent recidivism	-0.04 (0.09)	0.18	0.96 [0.80, 1.15]	.526
Non-violent recidivism	0.05 (0.06)	0.78	1.06 [0.94, 1.19]	.515

*Note.* \*\* $p < .01$ 

SE = standard error; HR = hazard ratio; CI = confidence interval

**Table 7***Cox Regression Results for Predicting Total Sexual Recidivism (N = 626)*

Covariates	B (SE)	Wald	HR [95% CI]	Harrell's C
Model	-	-	-	.642
Total SSPI-2 score	0.13 (0.10)	1.60	1.14 [0.93, 1.38]	-
Total Static-99R score	0.16 (0.05)	13.07***	1.18 [1.08, 1.29]	-

*Note.* \*\*\* $p < .001$ 

SE = standard error; HR = hazard ratio; CI = confidence interval

## Appendix A

### Open Science Framework Pre-Registration for Study 1

**Title**

The Screening Scale for Hebephilic Interests (SSHI)

**Description**

The current study is an analysis of archival data from a sexual behaviour clinic in Toronto, Canada. The dataset is comprised of offenders who were referred to the clinic for a sexological assessment and includes assessment information, such as victim information and phallometric assessment data for each offender. Victim information is comprised of official file information and official criminal history documentation. Phallometric testing is a psychophysiological measure of sexual arousal to different stimuli which vary by sex (male or female) and sexual development (prepubescent, pubescent, and adult). The purpose of this study is to examine victim characteristics that are positively associated with a primary sexual interest in pubescent children (hebephilia) and to combine these items to develop a scale that can accurately classify offenders with primary hebephilic interests from those who have no sexual interest in children. The classification of these two groups (primary hebephilia versus no sexual interest in children) is based on phallometric testing (see ‘Indices’).

**Category**

Project

**Affiliated institutions**

*Saint Mary’s University*

**License**

No license

**Year**

2020

**Copyright Holders**

N/A

**Subjects**

Psychology, Social and Behavioural Sciences, Social Psychology

## Tags

Assessment, Behavioural measure, Sex offenders, Sexual interest in children

## Study Information

### Hypotheses

**List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.**

1. It is hypothesized that two items from the Screening Scale for Pedophilic Interests-Revised (SSPI-2), will be positively associated with primary hebephilia. The two items include ‘multiple child victims under the age of 15’ and ‘child pornography offences.’

Exploratory analyses will be conducted to examine the remaining 3 items (‘any boy victims under the age of 15,’ ‘any child victims under the age of 12,’ and ‘any extrafamilial child victims under the age of 15’) on the SSPI-2 in their ability to classify offenders as having primary hebephilia. We have no hypotheses about these specific items.

2. We hypothesize that two additional items not on the SSPI-2 will be associated with primary hebephilia. These two items include ‘at least one victim between the ages of 15-16,’ and the ‘absence of gender polymorphism.’ Gender polymorphism refers to those who offend against both boys and girls without discriminating between sexes. Therefore, we hypothesize the absence of this, or those who offend exclusively against one gender will be positively associated with primary hebephilia.

3. Finally, we hypothesize that the items analyzed in hypothesis 1 and 2 will be able to be combined to develop a new measure (SSHI) to accurately classify those who have primarily hebephilic interests.

We will also examine the psychometric properties of the measure we develop, which include sensitivity, specificity, positive predictive value, and negative predictive value.

## Design Plan

### Study type

#### **Please check one of the following statements**

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, “natural experiments,” and regression discontinuity designs.

Meta-Analysis - A systematic review of published studies.

→ Other

### **Blinding**

**Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.**

→ No blinding is involved in this study.

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)

Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

**Is there any additional blinding in this study?**

**Blinding (Other)**

N/A

### **Study design**

**Describe your study design. The key is to be as detailed as is necessary given the specific parameters of the design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required.**

The design of the current study is an analysis of an archival clinical database with a between-subjects design. For the purpose of the study, the final sample (target and comparison groups) will be randomly split into two equal samples. The first group (development sample) will be used for the initial analyses (H1-H2) and creation of the measure (H3). The second group (validation sample) will be used for validation of the measure to ensure participants can be accurately classified as having primary hebephilia. Primary hebephilia is defined as participants having a much greater response to pubescent children relative to adults and having a higher response to pubescent children than

prepubescent children. See "Explanations of Existing Data" and "Data Collection Procedures" for further information on study design.

### Randomization

**If you are doing a randomized study, state how you will randomize, and at what level. Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.**

N/A

### Sampling Plan

#### Existing Data

**Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. See <https://cos.io/prereg> for more information.**

Registration prior to creation of data

Registration prior to any human observation of the data

Registration prior to accessing the data

Registration prior to analysis of the data

→ Registration following analysis of the data

#### Explanation of existing data

**If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.**

The current data comes from an archival database which includes both clinical and phallometric assessment data of men who underwent a sexological assessment at a sexual behaviour clinic in Toronto, Canada. The archival database was maintained at the clinic between 1995 and 2011.

The data has been previously used for prior research studies that have addressed distinct research questions. Of greatest relevance, is that the dataset was used for the development of the Screening Scale for Pedophilic Interests-Revised (SSPI-2; Seto et al., 2017; Stephens et al., 2019). The primary researcher was not involved in the development of the previously mentioned measure or any other study that has used this sample.

Although we have not tested our hypotheses, summary statistics have been conducted by the primary researcher in order to exclude certain participants from the study and to generate the target and comparison group (i.e., those who have primary hebephilia and those who have no sexual interest in children, respectively). Demographic information has also been analyzed for the purpose of writing and defending the thesis proposal. This descriptive information includes average age at assessment, average age at first and last known sexual offence, relationship status, ethnicity, referring agency, and psychiatric history. No other analyses have been tested using the target and comparison group and the hypotheses described above (including exploratory analyses) have not yet been tested.

### **Data collection procedures**

**Please describe the process by which you will collect your data and your inclusion and exclusion criteria. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool, and your study timeline. For studies that don't include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.**

The total archival database consisted of 3,343 men who had a charge for at least one sexual offence and underwent sexological assessment at a sexual behaviour clinic in Toronto between 1995 and 2011. To be included in the archival database, men had to consent to the use of their data for research purposes. Although the dataset includes several variables that were collected for the purposes of the original assessment (e.g., self-reported sexual interest, sexual partner information), only demographic data, victim characteristics, and phallometric test results will be used and analyzed for the present study. The entirety of the assessment information in the archival database was primarily collected by the phallometric technician, entered on a coding sheet and then into an ongoing clinical database by a single research assistant. Below we document the variables that are specifically being used in the present study.

**Victim Information:** Victim information was coded from official file information including prior assessments, official criminal history documentation (e.g., police reports) and through self-reported sexual history. Victim information recorded in the original database included the victim gender (male, female), age (0-5, 6-10, 11, 12-14, 15-16, 17+), and relationship the victim had to the offender (intrafamilial, extrafamilial). Victim counts were entered for each variable. When a discrepancy in victim count occurred between official file information and self-reported information, the source which had the highest number of victims was used.

**Phallometry:** Phallometric testing was conducted by a single phallometric technician. Briefly, participants were placed in a room separate from the technician and instructed to place a cuff at the base of their penis and a plastic cylinder over the cuff to create an

airtight seal. Changes in air displacement in the cylinder correspond to changes in penile blood volume in response to sexual stimuli.

Participants were shown different sets of still images and audio stories that varied based on sex (male or female) and sexual maturity (prepubescent, pubescent, or adult). The sets included: a) prepubescent girls (aged 3 to 11), b) prepubescent boys (aged 5 to 11), c) pubescent girls (aged 12 to 14), d) pubescent boys (aged 12 to 14), e) adult females (aged 20 to 34), and f) adult males (aged 19 to 41). The seventh category of stimuli was a neutral landscape image with an audio description and was presented as a control category. There were four trials per stimulus category for a total of twenty-eight trials. Each of the four trial scores for each stimulus category were averaged to create the seven category scores.

Because of the variability in participant characteristics (e.g., participant age, size of penis, time since last ejaculation, etc.), raw phallometric scores were transformed into z-scores for each trial and the scores were averaged for each stimulus category to create seven z-scores per stimulus set (Blanchard et al., 2001).

### **Sample size**

**Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, describe how many units are you collecting at each level of the analysis. This might be the number of samples or a range, minimum, or maximum.**

Individuals who had no victim information, and those who did not have child victims under the age of 15 were removed from the archival database. This created a total sample size of 1,900, which has been used for previous studies.

For the current study, we further excluded those who had a primary interest in prepubescent children (pedophilia;  $n = 61$ ) and those who were categorized as having pedohebephilia but responded higher to prepubescent than to pubescent children ( $n = 184$ ). The final sample size includes 1,655 individuals. The target group is those who have hebephilic interests (primary sexual interests in pubescent children;  $n = 567$ ) and the comparison group is those who have no sexual interests in children ( $n = 1,088$ ) based on phallometric testing.

### **Sample size rationale**

**This could include a power analysis or an arbitrary constraint such as time, money, or personnel.**

A power analysis was not conducted as all participants who met the inclusion criteria will be included in the analyses. This includes 1,655 participants with 1,088 of those

participants having no sexual interest in children (comparison group) and 567 having primary hebephilic interests (target group).

### **Stopping rule**

**If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection. If you are using sequential analysis, include your pre-specified thresholds.**

N/A

## **Variables**

### **Manipulated variables**

**Precisely define all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.**

N/A

### **Measured variables**

**Precisely define each variable that you will measure. This will include outcome measures, as well as any measured predictors or covariates.**

It is important to note that the dataset contains many other variables but only those which will be included in analyses will be described below. Four sets of variables will be described: variables used to create the target and comparison group (group variable), demographic variables, variables from the SSPI-2, and additional variables.

**1. Group Variable:** Phallometric z-scores for each stimulus category were used to create the target and comparison group. First, a pedophilia and hebephilia index score was created. This involved subtracting the maximum average score to adults, regardless of sex, from the maximum average score to prepubescent (for pedophilia index) or pubescent children (for hebephilia index). A cut-off score of  $z > 0.25$  on the indices were used to categorize offenders as having pedophilia or hebephilia.

After the phallometric indices were created they were combined, and a variable was created from the two indices to categorize all individuals into one of four different groups: those who met the pedophilic index only were considered “pedophilic only”, those who met the hebephilic index only were considered “hebephilic only”, those who met both indices were considered “pedohebephilic”, and those who did not meet either indices were considered to have “no sexual interest in children.”

Lastly, another variable was created to classify those who were considered “pedohebephilic” but responded greater to pubescent children than prepubescent children. Therefore, the target group will comprise those who met the hebephilic index only and those who were considered pedohebephilic and met both indices but had a greater

response to pubescent than prepubescent children. The comparison group will include those who did not meet either indices and were classified as having no sexual interest in children.

## **2. Demographic variables:**

- a) Ethnicity: categorized as Caucasian, African Canadian, East Indian/Pakistani, First Nations, Asian Canadian, Filipino/Pacific Islander and "other." Recorded during the clinical interview.
- b) Age at assessment: the age of the offender at the assessment at the clinic. Recorded during the clinical interview.
- c) Age at first known sexual offence: the age of the offender at their first known offence, based on official criminal record and self-report.
- d) Age at last known sexual offence: the age of the offender at their most recent or last known offence, based on official criminal record and self-report.
- e) Married or common-law: whether the offender had ever been married or lived common law with their partner.
- f) Referring agency: the agency that referred the offender to the clinic for assessment. Categorized as probation/parole, own lawyer, legal aid, institution, or self-referred.
- g) Psychiatric history: the psychiatric history of the offender assessed at the time of the clinical interview. Categorized as no known history, affective disorder, schizoaffective disorder, schizophrenia, or other.

## **3. SSPI-2 variables:**

- a) Any boy victims under the age of 15: whether the offender had at least one boy victim under the age of 15 (yes/no).
- b) Multiple child victims under 15: whether the offender had two or more child victims under the age of 15 (yes/no).
- c) Any child victims under the age of 12: whether the offender had at least one child victim under the age of 12 (yes/no).
- d) Any extrafamilial child victims under 15: whether the offender had at least one child victim unrelated to the offender. Extrafamilial is defined as a child who was not the offender's son or daughter, stepson or stepdaughter, nephew or niece, grandchild, or first cousin (yes/no).
- e) Any possession of child pornography: whether the offender had any charges or convictions for the possession of child pornography or if the offender self-admitted to possessing it (yes/no).

## **4. Additional variables:**

- a) Victims between 15-16: whether the offender had one or more victims between the ages of 15 and 16 (yes/no).

- b) Gender polymorphism: whether the individual had committed sexual offences against both girls and boys who were 15 years of age or younger (yes/no).

### Indices

**If applicable, please define how measures will be combined into an index (or even a mean) and what measures will be used. Include either a formula or a precise description of the method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), please note that here but describe the exact method in the analysis plan section.**

Using the phallometric z-scores, a pedophilic and hebephilic index were created by subtracting the individual's maximum average response to adults from their maximum average response to children (pubescent children for the hebephilia index and prepubescent children for the pedophilia index). A cut-off score of  $z > 0.25$  on the indices were used to categorize offenders as having pedophilia or hebephilia.

After the phallometric indices were created they were combined, and a variable was created from the two indices to categorize all individuals into one of four different groups. Those who met the pedophilic index only were considered "pedophilic only", those who met the hebephilic index only were considered "hebephilic only", those who met both indices were considered "pedohebephilic", and those who did not meet either indices were considered to have "no sexual interest in children." Another variable was created to classify those who were considered "pedohebephilic" but responded greater to pubescent children than prepubescent children. The target group will comprise those who met the hebephilic index only and those who were considered pedohebephilic and met both indices but had a greater response to pubescent than prepubescent children. The comparison group will include those who did not meet either indices and were classified as having no sexual interest in children.

## Analysis Plan

### Statistical models

**What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, RMANOVA, MANOVA, multiple regression, SEM, etc) and the specification of the model. This includes each variable that will be included, all interactions, subgroup analyses, pairwise or complex contrasts, and any follow-up tests from omnibus tests. If you plan on using any positive controls, negative controls, or manipulation checks you may mention that here. Provide enough detail so that another person could run the same analysis with the information provided. Remember that in your final article any test not included here must be noted as exploratory and that you must report the results of all tests.**

The target and comparison groups will be randomly divided into two subsamples and then recombined to create the development and validation groups. The target group will be divided in half to create two subsamples ( $n = 283$  and  $n = 284$ ) and the comparison

group will be divided the same way ( $n = 544$  and  $544$ ). One subsample from the target and one subsample from the comparison group will then be combined to create a development sample ( $n = 827$ ) and a validation sample ( $n = 828$ ) with an even number of participants coming from the target and comparison groups. The development sample will be used for the creation of the measure and the following three hypotheses will be conducted using this sample.

For the development sample ( $n = 827$ ):

1. To address the first hypotheses, regarding items from the SSPI-2, Chi-square analyses will be used to determine which individual items from the SSPI-2 will be positively associated with primary hebephilia.

Exploratory analyses using Chi-square analyses will also be used to examine the remainder of the 3 items on the SSPI-2 that were not part of the hypothesized items.

2. To address the second hypotheses, regarding additional items, Chi-square analyses will also be used to determine if the additional proposed items are positively associated with primary hebephilia.

3. Based on the first two hypotheses, we anticipate that we will be able to create a measure based on the significant items. Logistic regression will allow us to determine which items will be retained for the measure.

Once the measure has been created based on the logistic regression results, Area Under the Curve (AUC's) will be used to determine the ability of the new measure to classify those with primary hebephilia.

For the validation sample ( $n = 828$ ):

The second sample will be used for the validation of the new measure. Validation is important to ensure the measure is able to accurately classify the remainder of the sample as having hebephilia. AUC's will again be used for classification accuracy in determining the ability of the SSHI to correctly identify individuals with primary hebephilia.

For the total sample ( $n = 1,655$ ):

Lastly, descriptive statistics will also be conducted by running frequencies on the total sample. Demographic information includes average age at assessment, average age at first and last known sexual offence, married or common-law, ethnicity, referring agency, and psychiatric history. We will also examine the average z-score responses to each of the phallometric categories mentioned in "Measured variables." Finally, psychometric properties of the new measure (i.e., sensitivity, specificity, positive predictive value and negative predictive value) will also be examined on the total sample size. We will recombine both samples ( $n = 1,655$ ) and examine the psychometric properties using hand calculated frequencies.

### **Transformations**

**If you plan on transforming, centering, recoding the data, or requiring a coding scheme for categorical variables, please describe that process.**

N/A

### **Inference criteria**

**What criteria will you use to make inferences? Please describe the information you'll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?**

To determine whether specific victim characteristics are positively associated with hebephilia we will use Null Hypothesis Significance Testing ( $p < .05$ ). We will also examine effect sizes for all analyses.

### **Data exclusion**

**How will you determine which data points or samples if any to exclude from your analyses? How will outliers be handled? Will you use any awareness check?**

Specific exclusion criteria were discussed in "Sample size." To reiterate, they include the removal of individuals who did not have victim information or a child victim under 15 years old. Individuals who were categorized as "pedophilic only" and "pedohebephilic individuals who had higher responding to prepubescent children" will also be removed from analyses. No other individuals will be removed from the study.

### **Missing data**

**How will you deal with incomplete or missing data?**

N/A

### **Exploratory analysis**

**If you plan to explore your data to look for unspecified differences or relationships, you may include those plans here. If you list an exploratory test here, you are not obligated to report its results. But if you do report it you are obligated to describe it as an exploratory result.**

As mentioned in the hypotheses, exploratory analysis will be examined for three of the five SSPI-2 items ('any boy victims under the age of 15,' 'any child victims under the age of 12,' and 'any extrafamilial child victims under the age of 15'). We have no hypotheses about these three items and will report results as an exploratory result.

### **Other**

**If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related**

**work such as replications or work that uses the same data, or other helpful context would be appropriate here.**

Blanchard, R., Klassen, P., Dickey, R., Kuban. M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13*(1), 118-126. <https://doi.org/10.1037/1040-3590.13.1.118>

Seto, M.C., Stephens, S., Lalumière, M. L., & Cantor, J. M. (2017). The revised screening scale for pedophilic interests (SSPI–2): Development and criterion-related validation. *Sexual Abuse: A Journal of Research and Treatment, 29*(7), 619-635. <https://doi.org/10.1177/1079063215612444>

Stephens, S., Seto, M. C., Cantor, J.M., Lalumière, M.L. (2019). The screening scale for pedophilic interests-revised (SSPI-2) may be a measure of pedohebephilia. *The Journal of Sexual Medicine, 16*(10), 1655-1663. <https://doi.org/10.1016/j.jsxm.2019.07.015>

## Appendix B

### Open Science Framework Pre-Registration for Study 2

**Title**

Predictive Validity of the Revised Screening Scale for Pedophilic Interests (SSPI-2)

**Description**

The current study is an analysis of an archival recidivism dataset from a sexual behaviour clinic in Toronto, Canada. The dataset is comprised of offenders who were referred to the clinic for a sexological assessment and includes assessment data for each offender. Individuals in the database were scored on the Revised Screening Scale for Pedophilic Interests (SSPI-2). The measure is used to assess for pedohebephilia (i.e., a sexual interest in children) in offenders who have previously committed a sexual offence against a child. Individuals were also scored on the Static-99R, an established risk assessment used to determine their risk for sexual reoffending. Recidivism data was gathered from criminal records which were obtained from the Royal Canadian Mounted Police (RCMP). The purpose of this study is twofold. First, it is to examine the predictive validity of the SSPI-2 by investigating whether it can predict sexual recidivism in men who sexually offend against children. Second, the study will examine whether the SSPI-2 adds incremental validity to the Static-99R.

**Category**

Project

**Affiliated institutions**

*Saint Mary's University*

**License**

No license

**Year**

2021

**Copyright Holders**

N/A

**Subjects**

Psychology, Social and Behavioural Sciences, Social Psychology

**Tags**

Predictive validity, Recidivism, Incremental validity, Child sexual offenders

## Study Information

### Hypotheses

**List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.**

1. Those who score higher on the SSPI-2 will be more likely to sexually recidivate than those who score lower on the measure.
2. Those who score higher on the SSPI-2 will be no more likely to commit violent and non-violent recidivism offences than those who score lower on the measure.
3. The SSPI-2 will **not add** to the Static-99R in the prediction of sexual recidivism.

### Design Plan

#### Study type

**Please check one of the following statements**

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, “natural experiments,” and regression discontinuity designs.

Meta-Analysis - A systematic review of published studies.

→ Other (Archival data)

#### Blinding

**Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.**

→ No blinding is involved in this study.

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)

Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

**Is there any additional blinding in this study?**

**Blinding (Other)**

N/A

**Study design**

**Describe your study design. The key is to be as detailed as is necessary given the specific parameters of the design. There may be some overlap between this question and the following questions. That is OK, as long as sufficient detail is given in one of the areas to provide all of the requested information. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required.**

This study will use an archival database of individuals who were assessed at a sexual behaviour clinic and had recidivism data collected as part of a separate project. All participants that are eligible for inclusion and had complete recidivism data coded will be used in each of the analyses. See "Explanations of Existing Data" and "Data Collection Procedures" for further information on study design.

**Randomization**

**If you are doing a randomized study, state how you will randomize, and at what level. Typical randomization techniques include: simple, block, stratified, and adaptive covariate randomization. If randomization is required for the study, the method should be specified here, not simply the source of random numbers.**

N/A

### **Sampling Plan**

**Existing Data**

**Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. See <https://cos.io/prereg> for more information.**

Registration prior to creation of data

Registration prior to any human observation of the data

Registration prior to accessing the data

**→ Registration prior to analysis of the data**

Registration following analysis of the data

### **Explanation of existing data**

**If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study.**

The current archival dataset includes recidivism information, sexological assessment information (i.e., phallometric results, victim information), and risk assessment information. The sample is comprised of men who underwent a sexological assessment at a sexual behaviour clinic in Toronto, Canada between 1995 and 2006. The recidivism information is inclusive of reoffending up until 2013.

The recidivism data has been previously used for three prior research studies that have addressed distinct research questions (e.g., Stephens et al., 2017). The primary researcher was not involved in any of the previous studies that have used this sample. Although none of the hypotheses have been tested, demographic information has been analyzed for the purpose of writing and defending the thesis proposal. This descriptive information includes average age at assessment, average age at first and last known sexual offence, relationship status, ethnicity, referring agency, and psychiatric history. No other analyses have been conducted.

### **Data collection procedures**

**Please describe the process by which you will collect your data and your inclusion and exclusion criteria. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool, and your study timeline. For studies that don't include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.**

Men who were assessed at CAMH between 1995 and 2011 ( $n = 3,343$ ) were included in the recidivism database if they had complete assessment information, valid phallometric testing assessed after 1995 but before 2006 to ensure adequate follow-up time, at least one sexual assault victim of any age, and if they were 18 or older at the time of the assessment.

For the recidivism data, 1,117 criminal records were ordered from the Royal Canadian Mounted Police (RCMP) based on the above inclusion criteria, but only 844 were obtained, as 273 of the records were not accessible from the national database. From the 844 criminal records, 626 were included for analyses in the current study based on additional inclusion criteria: the offender had at least one child victim under the age of 15; they were not in custody during the entire follow-up time and thus had the opportunity to reoffend in the community; and they had complete data for the total SSPI-2 and Static-99R scores.

Although the dataset includes several variables that were collected for the purposes of the original assessment (e.g., self-reported sexual interest, sexual partner information), only demographic variables, recidivism data, opportunity to reoffend, and total scores on the SSPI-2 and Static-99R will be used in the present study. The original assessment information in the database was primarily collected by a phallometric technician, entered on a coding sheet and then into an ongoing clinical database by a single research assistant. The recidivism data (i.e., recidivism, opportunity to reoffend, and the Static-99 scores) were coded by Skye Stephens and research assistant, Kylie Reale, based on the criminal record sent by the RCMP and the phallometric reports.

### **Sample size**

**Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, describe how many units are you collecting at each level of the analysis. This might be the number of samples or a range, minimum, or maximum.**

The original sample from the archival database was comprised of 3,343 men. Since this study is based on recidivism, criminal records were ordered for 1,117 men. Only 844 were obtained from the RCMP and 626 offenders will be included for analyses based on the above-described inclusion criteria (under Data Collection Procedures).

### **Sample size rationale**

**This could include a power analysis or an arbitrary constraint such as time, money, or personnel.**

A power analysis was not conducted for this study. Recidivism data was only ordered for a subset of cases from the larger assessment database that had complete assessment information, valid phallometric testing after 1995 but before 2006 to ensure adequate follow-up time, at least one sexual assault victim of any age, and were 18 or older at the time of assessment. The number of recidivism cases ordered was based on the budget allotted for Dr. Skye Stephen's Ph.D. dissertation and therefore, files could not be ordered for every case in the archival database. The analyses will be conducted on all available cases that meet the following inclusion criteria: the offender had at least one child victim under the age of 15; they were not in custody during the entire follow-up time and thus had the opportunity to reoffend in the community; and they had complete data for the total SSPI-2 and Static-99R scores.

### **Stopping rule**

**If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection. If you are using sequential analysis, include your pre-specified thresholds.**

N/A

## Variables

### Manipulated variables

**Precisely define all variables you plan to manipulate and the levels or treatment arms of each variable. This is not applicable to any observational study.**

N/A

### Measured variables

**Precisely define each variable that you will measure. This will include outcome measures, as well as any measured predictors or covariates.**

1. **Opportunity to reoffend (time to event):** Opportunity to reoffend was operationally defined as the amount of time an offender resided in the community during the follow-up period, while accounting for time spent in secure custody. Participants who did not have an opportunity to reoffend were removed from analyses (i.e., individuals who were incarcerated for the entire follow up period). Cases that were censored, e.g., the participant died, were followed to that end-point. To calculate opportunity to reoffend, a composite score was calculated that involved subtracting the total number of months spent in secure custody since the start of the follow-up period from each offender's total follow-up time. This allowed for a more accurate calculation of the opportunity each offender had to reoffend in the community because it considered time spent in secure custody as opposed to total follow-up time. Opportunity to reoffend was considered for each recidivism outcome, as it could vary by recidivism type. Three variables for opportunity to offend will be used in the present study which include the following:
  - a) Total sexual opportunity to reoffend
  - b) Violent opportunity to reoffend
  - c) Non-violent opportunity to reoffend
2. **Recidivism (event):** Data for recidivism was collected through the Canadian Police Information Centre (CPIC) of the RCMP, a Canadian federal policing agency. CPIC documents included a federal identifying number for each individual, criminal charges and convictions, and their dispositions for crimes that occurred anywhere in Canada. Recidivism was considered to be any charges or convictions for new offences that occurred after the initial sexological assessment. The data also includes the exact dates of new charges/convictions of recidivism for each offender. This is important because Cox regression analyses use "survival time" or the time that it takes from the point of origin (i.e., after the assessment) until the event of interest occurs (i.e., recidivism). To capture the most accurate rate of reoffending both charges and convictions were considered in

this study. Using the CPIC documents, recidivism was gathered by counting new convictions and charges and creating composite variables. Recidivism in the original database was coded into four different categories: contact sexual offences (e.g., sexual assault), non-contact sexual offences (e.g., voyeurism), nonsexually violent offences (e.g., assault), and non-violent offences (e.g., condition breaches). For this study, contact and non-contact sexual offences will be combined together and analyzed as total sexual recidivism. Violent and non-violent recidivism will be examined separately. All recidivism variables were coded dichotomously with '1' indicating one or more recidivism offences occurred and '0' indicating no recidivism occurred during the opportunity to reoffend.

- a) Total sexual recidivism included charges and convictions for contact and non-contact sexual offences.
- b) Violent recidivism included charges and convictions for non-sexual violent offences, such as general or aggravated assaults
- c) Non-violent recidivism included charges and convictions for non-sexual, non-violent offences, such as conditional breaches.

- 3. Total SSPI-2 Score (predictor variable):** The SSPI-2 is a measure of pedohebephilic interests among male offenders who have previously committed a sexual offence against a child (Seto et al., 2017). It consists of five items including: any boy victim under the age of 15, multiple child victims under the age of 15, any child victims under the age of 12, any extrafamilial child victims under the age of 15, and any charges for child pornography offences. Scores on the SSPI-2 can range from zero to five, with higher scores indicating a greater likelihood of pedohebephilia. Items from the SSPI-2 were scored based on victim characteristics from self-report combined with file documentation. Only total SSPI-2 scores will be used in this study.
- 4. Total Static-99R Score (predictor variable):** The Static-99R is an actuarial risk assessment used to determine risk for sexual reoffending in adult male sex offenders (Hanson & Thornton, 2000). The measure contains 10 items. Eight out of the ten items are scored dichotomously and include: ever lived with an intimate partner, conviction for non-sexual violence at the time of index, any convictions for non-sexual violence prior to the index offence, four or more prior sentencing dates, convictions for non-contact sexual offences, any unrelated victims, any stranger victims, and any male victims. The remaining two items are scored differently. Age at release is scored from 1 to -3, the younger the individual being the higher the score. The last item, prior charges and convictions for sexual offences is scored from 0 to 3, the higher the score indicating more charges and convictions. Total Static-99R scores can range from -3 (indicating low risk of reoffending) to 11 (indicating a high risk of reoffending). Only total Static-99R scores will be used in this study.

**5. Demographic variables:** demographic variables will be used to describe the sample and include the following variables:

- h) Ethnicity (as recorded in the original database): categorized as Caucasian, African Canadian, East Indian/Pakistani, First Nations, Asian Canadian, Filipino/Pacific Islander and "other." Recorded during the clinical interview.
- i) Age at assessment (as recorded in the original database): the age of the offender at the assessment at the clinic, in years. Recorded during the clinical interview.
- j) Age at first known sexual offence (as recorded in the original database): the age of the offender at their first known offence, in years, based on official criminal record or self-report.
- k) Age at last known sexual offence (as recorded in the original database): the age of the offender at their most recent or last known offence, based on official criminal record and self-report.
- l) Married or common-law (as recorded in the original database): whether the offender had ever been married or lived common law with their partner.
- m) Referring agency (as recorded in the original database): the agency that referred the offender to the clinic for assessment. Categorized as probation/parole, own lawyer, legal aid, institution, or self-referred.
- n) Psychiatric diagnosis (as recorded in the original database): the primary psychiatric diagnosis of the offender assessed at the time of the clinical interview. Categorized as no known diagnoses, affective disorder, schizoaffective disorder, schizophrenia, or other.

## Indices

**If applicable, please define how measures will be combined into an index (or even a mean) and what measures will be used. Include either a formula or a precise description of the method. If you are using a more complicated statistical method to combine measures (e.g. a factor analysis), please note that here but describe the exact method in the analysis plan section.**

N/A

## Analysis Plan

### Statistical models

**What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, RMANOVA, MANOVA, multiple regression, SEM, etc) and the specification of the model. This includes each variable that will be included, all interactions, subgroup analyses, pairwise or complex contrasts, and any follow-up tests from omnibus tests. If you plan on using any positive controls, negative**

**controls, or manipulation checks you may mention that here. Provide enough detail so that another person could run the same analysis with the information provided. Remember that in your final article any test not included here must be noted as exploratory and that you must report the results of all tests.**

Descriptive statistics will be conducted by running frequencies on the total sample. Demographic information includes average age at assessment, average age at first and last known sexual offence, married or common-law, ethnicity, referring agency, and psychiatric diagnosis.

To examine the first two hypotheses (the predictive validity of the SSPI-2), Cox regressions will be used, and Harrell's concordance index (Harrell's C) will be calculated as an effect size estimate using R statistical software. Cox regressions will be conducted for each recidivism outcome (i.e., sexual, nonsexually violent, and nonviolent recidivism). To examine the third hypothesis, whether the SSPI-2 adds incremental validity to the Static-99R, a cox regression will also be used. As the Static-99R is used to determine risk of sexual reoffending, this analysis will use only the sexual recidivism outcome. Based on research by Babchishin et al. (2012) which found that large sample sizes typically result in incremental validity with highly correlated measures, both the SSPI-2 and Static-99R will be entered together in the same block of the cox regression.

### **Transformations**

**If you plan on transforming, centering, recoding the data, or requiring a coding scheme for categorical variables, please describe that process.**

N/A

### **Inference criteria**

**What criteria will you use to make inferences? Please describe the information you'll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?**

To examine whether higher SSPI-2 scores are associated with sexual recidivism and whether the SSPI-2 adds incremental validity to the Static-99R, we will use Null Hypothesis Significance Testing ( $p < .05$ ). Harrell's concordance index (Harrell's C) will also be calculated using R statistical software as an effect size. Harrell's C is used as an alternative to AUCs when follow-up time is different for each offender, which is the case for this study (Helmus & Babchishin, 2017).

### **Data exclusion**

**How will you determine which data points or samples if any to exclude from your analyses? How will outliers be handled? Will you use any awareness check?**

Specific inclusion criteria were discussed above under “Sample Size.” To reiterate, offenders who did not have at least one child victim under the age of 15, did not have the opportunity to reoffend in the community, and were missing data for the Static-99R and the SSPI-2, were also removed. No other individuals will be removed from the study.

### **Missing data**

#### **How will you deal with incomplete or missing data?**

Cases with missing data for the total SSPI-2 or Static-99R scores will be removed prior to analyses.

### **Exploratory analysis**

**If you plan to explore your data to look for unspecified differences or relationships, you may include those plans here. If you list an exploratory test here, you are not obligated to report its results. But if you do report it you are obligated to describe it as an exploratory result.**

N/A

### **Other**

**If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other helpful context would be appropriate here.**

Babchishin, K., Hanson, R., & Helmus, L. (2012). Even highly correlated measures can add incrementally to predicting recidivism among sex offenders. *Assessment, 19*(4), 442-461. <https://doi.org/10.1177/1073191112458312>

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119–136. <https://doi.org/10.1023/A:1005482921333>

Helmus, L., & Babchishin, K. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*(1), 8-25. <https://doi.org/10.1177/0093854816678898>

Seto, M.C., Stephens, S., Lalumière, M. L, & Cantor, J. M. (2017). The revised screening scale for pedophilic interests (SSPI-2): Development and criterion-related validation. *Sexual Abuse: A Journal of Research and Treatment*, 29(7), 619-635.  
<https://doi.org/10.1177/1079063215612444>