CONCEPTUAL AND EMPIRICAL ISSUES IN PEDOHEBEPHILIC INTEREST

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in the Department of Psychology

University of Saskatchewan

Saskatoon, Saskatchewan

By

Ian V. McPhail

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department of the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Psychology
9 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5A5,
Canada

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan, S7N5C9,
Canada

ABSTRACT

Male sexual attraction to children is central for understanding and preventing sexual offending against children. The current research aimed to examine conceptual and empirical issues in how we understand, treat, and measure pedohebephilic interests in men with histories of sexual offences. Chapter 2 and 3 presents a taxometric analysis of phallometric measures of pedophilic interest. The results of the taxometric analysis indicated that pedophilic interest is distributed as three latent categories. In post-hoc analyses, the three categories were characterized as having non-pedophilic, non-preferentially pedophilic, and preferentially pedophilic interest. Chapter 4 presents a meta-analysis of research examining the effect of interventions in reducing pedophilic arousal in men with histories of sexual offences against children. The results of the meta-analysis suggest that behavioural and pharmacological interventions were associated with significant decreases in sexual arousal. Some of these interventions showed magnitudes of change that were greater than estimated natural history processes and men in some of these interventions showed posttreatment levels of arousal comparable to non-offending men. Men with the highest levels of pretreatment pedophilic arousal demonstrated the greatest amount of change over the course of treatment. In contrast, no intervention approaches were associated with increases in arousal to adults. Chapter 5 examined convergent and predictive validity in three measures of pedophilic interest. The results indicate that a phallometric test and the Violence Risk Scale-Sexual Offense version's (VRS-SO) Sexual Deviance factor showed convergence, while the Screening Scale for Pedophilic Interest (SSPI) demonstrated somewhat less convergence with the other measures. Similarly, the phallometric test and the Sexual Deviance factor of the VRS-SO were predictive of sexual recidivism and the VRS-SO remained a significant predictor after controlling for static risk. The SSPI was not predictive of sexual recidivism. Within the phallometric test, the

predictive validity of different latent structural models of pedophilic interest were examined. Across the models, a dimensional model and a trichotomous model received the most support, with the latter remaining predictive of sexual recidivism after controlling for static risk. Taken together, the results identify the need for replication studies examining latent structure in pedophilic interest, demonstrate the potential for treatment to help improve men's ability to regulate their sexual arousal, and provide further validity evidence for measures of pedophilic interest.

ACKNOWLEDGEMENTS

As with any complex and lengthy undertaking, a lot of people contribute to a dissertation coming alive and reaching completion. I would first like to thank my supervisor Dr. Mark Olver for his expert advice and guidance through the ups and downs of these projects. My projects relied on the work of others, and I am grateful to Drs. Sebastien Brouillette-Alarie and Jan Looman for allowing access to data for our taxometric analysis. I would also like to thank the members of my dissertation committee, Drs. Sarah Hoffman, Terry Nicholaichuk, and Steve Wormith for their help and input into these projects as the research unfolded.

I am also grateful to the Social Sciences and Humanities Research Council for their financial support which allowed me to focus my energy on doing this research. I am also grateful to the Centre for Forensic Behavioural Sciences and Justice Studies at the University of Saskatchewan and Public Safety Canada for providing research grant monies in support of the meta-analysis in my dissertation. Stephanie Fernane helped code articles for the meta-analysis, and I am indebted to her for bringing this project over the goal line.

Last, I would like to acknowledge the support of my family, partner, and friends. Having personal support and people around to listen to my ideas, worries, and joys as this research unfolded has made the journey all the more sweet.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Appendix D. Benchmark comparisons

LIST OF APPENDICES

CHAPTER 1

GENERAL INTRODUCTION

Pedohebephilic interest, which connotes sexual attractions to prepubescent and/or early pubescent children, is a central construct in human sexuality and for understanding why some men pursue sexual contacts with children. Theories of sexual offending against children include pedohebephilic interest as a key construct that motivates some men to commit sexual offences (Finkelhor & Araji, 1986; Hall & Hirshman, 1992; Seto, 2018; Ward & Beech, 2006; Ward & Seigert, 2002). Empirical studies show that men who commit sexual offences against children display, on average, higher levels of pedohebephilic interests across multiple measurement modalities (Implicit Association Tests, Babchishin, Hermann, & Nunes, 2013; phallometric testing, McPhail et al., 2017; viewing time measures, Schmidt, Babchishin, & Lehmann, 2017). Theory also posits that pedohebephilic interests are central to understanding why some men continue to commit sexual offences after receiving a criminal justice sanction for this behaviour (Mann, Hanson, & Thornton, 2010; Ward & Beech, 2004). Empirical studies support the status of pedohebephilic interest as a risk factor for future sexual violence (Hanson & Morton-Bourgon, 2005; McPhail et al., 2017). The available evidence indicates that pedohebephilic interest is central to understanding sexual offences against children.

1.1 Latent Structure in Pedohebephilic Interest

All psychological constructs are presumed to have a latent structure. Latent structure refers to the nature of how a psychological construct is distributed in the population (Meehl, 2004; Ruscio, Haslam, & Ruscio, 2006). One of the main differences within latent structure of psychological constructs is between latent dimensions and latent categories. Psychological constructs that are a latent dimension are distributed to varying degrees. Each individual in the

population has some standing on the psychological construct, with some individuals exhibiting very little of the construct and others exhibiting a great deal of the construct. For psychological constructs that are distributed as latent categories, the construct will be present for some individuals and not present for others. Latent categorical structure suggests that individuals are different from each other in type, not in degree.

The importance of latent structure to psychological science is hard to understate. Understanding latent structure has consequences to all downstream uses of a psychological construct: selecting research designs and statistical analyses, identifying individuals' standing on a construct in clinical assessments, and identifying individuals who may require treatment are decisions that depend on a presumed latent structure (Ruscio & Ruscio, 2004). Ideally, such applied and research decisions are based on empirical investigations of latent structure.

When empirically investigated, most psychological and psychiatric constructs have a dimensional structure, while few have been found to have categorical structure (Haslam, Holland, & Kuppens, 2012). The latent structure of pedophilic interest is poorly understood at present. Theoretical models identify pedophilic interest variously as having a categorical latent structure (Hanson, 2010; Seto, 2017) or dimensional latent structure (Finkelhor & Araji, 1986). Categorical models suggest that among the adult male population, there are men who exhibit pedophilic interest (i.e., pedophilic men) and men who exhibit no pedophilic interest (i.e., non-pedophilic men). Dimensional models suggest that all adult men exhibit pedophilic interest to some extent, with differences between men being differences in degrees of pedophilic interest.

Divergent theoretical models necessitate empirical investigation to further the understanding of latent structure in pedophilic interest. Recent studies have found support for categorical and dimensional latent structure in pedophilic interest (Schmidt, Mokros, & Banse,

2013; Stephens, Leroux, Skilling, Cantor, & Seto, 2017). These studies are meaningful and valid examinations of latent structure; however, given the conflicting results, the amount of empirical support for one latent structure over the other is equivocal.

1.2 Interventions for Pedohebephilic Interest

Constructs thought to be putative causal factors for sexual offending and risk factors for sexual recidivism are central to interventions aiming to reduce sexual recidivism. These constructs have generally been shown to predict future sexual recidivism in men with histories of sexual offending and are conceptualized as *criminogenic needs* (Bonta & Andrews, 2017) and *psychologically meaningful risk factors* (Mann et al., 2010). Pedohebephilic interests has been found to predict sexual recidivism in past research (k = 16, N = 1,961, d = 0.44; McPhail et al., 2019). Given this predictive relationship and the centrality of pedohebephilic interest in theoretical models of sexual offending against children, a majority of sexual offence treatment programs in North America aim to improve men's ability to control their arousal to children (McGrath et al., 2010).

Several interventions for pedohebephilic interest are based on behavioural principles. Behavioural interventions based on principles of punishment aim to improve men's ability to control their arousal to children using aversive procedures. These procedures can involve pairing fantasy about children or masturbation to child fantasies with a physically aversive stimulus, such as a foul odour, or a mentally aversive stimulus, such as imagining being discovered by the police while having sexual contact with a child (Marshall, Anderson, & Fernandez, 1999). Behaviour interventions based on the principles of reward pair an inherently rewarding stimulus, such as orgasm, with fantasy about adult individuals. For instance, orgasmic reconditioning procedures involve a male masturbating to a fantasy involving a child and, in the moments,

leading to orgasm, switching to a fantasy involving an adult (Marshall et al., 1999). Behavioural interventions can also be based on the principles of extinction, which involve masturbating to orgasm and continuing to masturbate during the post-orgasm refractory period and imagining sexual contact with a child (Marshall et al., 1999).

Pharmacological interventions of pedohebephilic interests have also been developed and are provided to men with sexual offence histories who display a pattern of arousal relevant to re-offending (i.e., pedophilic arousal; Association for the Treatment of Sexual Abusers [ATSA], 2012). Pharmacological interventions are generally antiandrogens, which reduce testosterone, and have an antilibidinal effect (Prentky, 1997). Some of the most commonly used pharmacological interventions used with men convicted of sexual offences are medroxyprogesterone acetate (MPA), cyproterone acetate (CPA), and leuprolide acetate (ATSA, 2012; Briken & Kafka, 2007; Prentky, 1997).

Other treatment approaches target pedohebephilic interests within a comprehensive treatment program that targets a variety of psychosocial issues. Such interventions may include behavioural interventions as a component that is provided to improve arousal control, while other modules of such programs can include improving intimacy skills and psychoeducation about healthy sexuality (Marshall et al., 1999).

Reviews of behavioural treatments tend to identify such procedures as leading to improved ability to manage arousal to children and less self-reported fantasy involving children in men with histories of sexual offending (Kelly, 1982; Turner & Briken, 2018). To date, no systematic and quantitative review has been undertaken to establish the effectiveness of various intervention modalities. Establishing the overall effectiveness of interventions for pedohebephilic

interest remains an important and outstanding issue in clinical approaches to men with these interests.

1.3 Validity in Measures of Pedophilic Interest

Putative psychologically meaningful risk factors should be included in clinical evaluations of risk for sexual recidivism. Outstanding issues remain regarding the clinical assessment of pedophilic interest and the status of pedophilic interest as a risk factor for future sexual offending. First, multiple measures exist to assess for the presence of pedophilic interest and measures used in clinical practice require ongoing validation research. Such measures include psychophysiological measures (e.g., phallometric testing), interview and file-based actuarial measures (e.g., the Sexual Deviance Factor of the Violence Risk Scale-Sexual Offense version [VRS-SO]; Wong, Olver, Nicholaichuk, & Gordon, 2003, 2017), and file-based measures (Screening Scale of Pedophilic Interest [SSPI]; Seto & Lalumière, 2001). While there is varying degrees of evidence to support these measures as predictive of sexual recidivism (Beggs & Grace, 2010; Helmus, Ó Ciardha, & Seto, 2014; McPhail et al., 2017; Olver et al., 2007; Seto, Sandler, & Freeman, 2017), there is relatively little evidence examining the whether these measures assess similar or distinct constructs. Past research has found that the VRS-SO Sexual Deviance factor, phallometric testing, and the SSPI are moderately correlated with each other (Canales, Olver, & Wong, 2009). Other research finds that the SSPI has a significant, yet small relationship with phallometric testing (Seto & Lalumière, 2001). Much of the available research has not, however, examined whether measures of pedophilic interest add meaningfully to the prediction of sexual recidivism after static risk factors are accounted for. Presumably, if pedophilic interest does not add to the identification of higher risk in men with histories of sexual offences, this is evidence against including it in clinical evaluations of risk.

Another issue is the extent to which latent structural models perform optimally in research examining the predictive validity of measures of pedophilic interest. Framing the past research in the language of *latent structural models*, the available evidence has examined whether dimensional, dichotomous, or trichotomous models of pedophilic interest predict sexual recidivism. Most studies included in a recent meta-analysismeasured pedophilic interest dimensionally and support this operationalization as predicting sexual recidivism (McPhail et al., 2019). However, a recent study with a large sample of men found that a dimensional model did not predict sexual recidivism (Stephens, Cantor, Goodwill, & Seto, 2017). Interestingly, most studies that have operationalized pedophilic interests dichotomously, by grouping samples into pedophilic and non-pedophilic men, have found that pedophilic interest is unrelated to sexual recidivism (Eher, Rettenberger, Matthes, & Schilling, 2010; Moulden et al., 2009; Stephens et al., 2017; Wilson, Abracen, Looman, Pichea, & Ferguson, 2010). This finding is contrary to other research and theoretical expectations, as men who exhibit sexual attractions to children are predicted to re-offend at a higher rate than men without such interests (Mann et al., 2010; Ward & Beech, 2004). This raises questions regarding status of pedophilic interest as a psychologically meaningful risk factor for sexual recidivism.

One potential explanation for the variation in findings is that pedophilic interest does not have a dichotomous latent structure and modelling the construct in such is suboptimal for statistical analyses. A related explanation is psychometric: the cut-offs used to identify pedophilic individuals were not optimal. A second potential explanation is that a dimensional latent structure best characterizes pedophilic interest, and research designs and statistical models that model pedophilic interest as dimensional are optimal.

A third possibility is that pedophilic interest has latent structure not captured by either model. Within a separate model, there are three categories of men: those who are not pedophilic, those who are non-preferentially or non-exclusively pedophilic, and those who are preferentially or exclusively pedophilic. This operationalization is currently used in diagnostic nosology (i.e., the 5th Diagnostic and Statistical Manual of the American Psychiatric Association), which includes a specifier to identify individuals who are exclusively attracted to children. The separation between preferentially or exclusively pedophilic men and non-preferentially pedophilic men may be taxonic. However, it is also possible that exclusive and preferential men exist on the higher end of a pedophilic dimension. Such a test remains to be conducted.

Research has used this exclusivity specifier and has found that those men who are exclusively pedophilic have high rates of sexual recidivism (Eher et al., 2010; Eher et al., 2015; Beier, 1998). Other research that has used phallometric testing to distinguish between non-preferentially and preferentially pedophilic men has found higher rates of sexual recidivism in the preferentially pedophilic group (Wilson et al., 2011). This research raises the possibility that preferential or exclusive pedophilic interest is supported as a risk factor for sexual recidivism, while non-preferential pedophilic interest is not a risk factor. The conflicting results and the possibility of using different latent structural models to operationalize pedophilic interest points to a need for further empirical examination. A related possibility that can explain the above-mentioned findings is that men with preferential or exclusive pedophilic interest are more likely to experience other risk factors for sexual recidivism (e.g., inability or no desire to establish intimate relationships with adults; Mann et al., 2010). The presence of these additional risk factors may explain the elevated rates of sexual recidivism by these men.

1.4 Conclusion

The present dissertation examines the latent structure of pedophilic interest to identify which latent structure may best characterize this sexual interest. The results of the latent structural findings will inform the next two studies in the dissertation, which are a meta-analysis of the existing research on the effectiveness of interventions for pedophilic interest and an examination of validity in three measures of pedophilic interest.

CHAPTER 2

TAXOMETRIC ANALYSIS OF THE LATENT STRUCTURE OF PEDOPHILIC INTEREST

This chapter has been previously published:

McPhail, I. V., Olver, M. E., Brouillette-Alarie, S., & Looman, J. (2018). Taxometric analysis of the latent structure of pedophilic interest. *Archives of Sexual Behavior, 47,* 2223–2240. doi:10.1007/s10508-018-1225-4

2.1 Introduction

Understanding the latent structure of psychological constructs requires empirical investigation (Beauchaine, 2003; Meehl, 1992, 1995a). Latent structure refers to the unobservable nature of a construct that is estimated using observable measurements on psychological tests. Two ways to conceptualize latent structure is as a single distribution or as two or more distinct classes (i.e., dimensional or taxonic; Meehl, 2004). Taxometric analyses are a family of analytic procedures that test whether the latent structure of a construct is best characterized as taxonic or dimensional (Ruscio, Haslam, & Ruscio, 2006; Waller & Meehl, 1998). The use of taxometric analyses has improved the understanding of the latent structure of a wide variety of psychological constructs (e.g., anxiety disorders, eating disorders, personality disorders; Haslam, Holland, & Kuppens, 2012).

Pedophilia, a sexual interest in prepubescent children, is an important construct in understanding, predicting, and preventing sexual offending against children (McPhail et al., 2017; Seto, 2018). Theoretical models predict that while most paraphilias in men are dimensional, pedophilic interest is likely to be taxonic and that men are either pedophilic or non-

pedophilic (Hanson, 2010). The prediction for taxonic structure is based on the early onset of pedophilic interest and high stability over the life course (Bailey, Hsu, & Bernhard, 2016; Hanson, 2010; McPhail, 2018; Seto, 2012; Tozdan & Briken, 2015). Gradient models of erotic age interests can be though to conceptualize pedophilic interest as a dimension (Blanchard, Kuban, Blak, Klassen, Dickey, & Cantor, 2012). However, gradient models can likely accommodate a finding that sexual interest in prepubescent children is taxonic, as most teleiophilic men may experience interest in physically mature adolescents, but little to no sexual interest in prepubescent children (Seto, 2017). Seto (2017) makes this hypothesis explicit, suggesting that it would be rare for an individual to have rather contrasting chronophilic attraction. Other theoretical work has conceptualized pedophilic interest as dimensional, suggesting men range from low levels of pedophilic interests to high levels of pedophilic interests (Finkelhor & Araji, 1986).

Without a robust empirical understanding of latent structure, researchers and clinicians must rely on untested assumptions regarding the structure of pedophilic interest to inform research and practice. For instance, past research examining neurological and neurodevelopmental correlates of pedophilic interest assume these interests are taxonic (Cantor & Blanchard, 2012; Cantor et al., 2008; Cantor et al., 2015; Dyshniku, Murray, Fazio, Lykins, & Cantor, 2015; Fazio, Dyshniku, Lykins, & Cantor, 2015; McPhail & Cantor, 2015). Certain clinical assessment procedures using phallometric testing also model pedophilic interest as dichotomous (i.e., two-ordered taxa; Blanchard, Klassen, Dickey, Kuban, & Blak, 2001; Cantor & McPhail, 2015), while other procedures assess pedophilic interest as a dimension (Marshall, O'Brien, & Marshall, 2009). The Pedophilic Disorder diagnostic criteria in the Diagnostic and Statistical Manual for Mental Disorders (DSM-5) represent an assumption of taxonic, but

trichotomous structure[1] to the disorder (i.e., three-ordered taxa; American Psychiatric Association [APA], 2013). Research examining treatment outcomes assume measurement models of pedophilic interest that are dichotomous (Müller, Curry, Ranger, Briken, Bradford, & Fedoroff, 2014) or dimensional (Becker, Kaplan, & Kavoussi, 1988; Bradford & Pawlak, 1993; Marques, Nelson, West, & Day, 1994; Marshall, 1997; Ricci, Clayton, & Shapiro, 2006). Problems may arise when latent structure does not align with the measurement model used in research, assessment, and treatment due to the use of less than optimal research design, statistical procedures, diagnostic classifications, and treatment expectancies (Ruscio et al., 2006).

2.1.1 Taxometric Research on Pedophilic Interest

Schmidt and colleagues (2013) conducted a taxometric analysis of pedophilic interest using Implicit Association Test (IAT; Greenwald, Schwartz, & McGee, 1998), viewing time, and self-report measures of pedophilic interest. Their results supported a taxonic structure in pedophilic interest. Those authors also conducted a latent profile analysis, which found a two-class solution and a three-class solution modelled their data equally well. These findings raise the possibility that a dichotomous or trichotomous latent structure may characterize pedophilic interest. Alternatively, if there are three groups, a dimensional structure of increasing severity may characterize pedophilic interest. Other research examined the latent structure of pedophilic interests using phallometric, self-report, and behavioural measures in a large sample of sexual offenders (Stephens, Leroux, Skilling, Cantor, & Seto, 2017). Those investigators found support for dimensional structure in pedophilic interest. A third taxometric study examined latent

[1] Using DSM-5 criteria, clinicians make a first categorical decision, whether a client has Pedophilic Disorder or does not have the disorder. Using the exclusivity specifier, clinicians make a second decision, whether a client diagnosed with Pedophilic Disorder is exclusively or non-exclusively attracted to children. This results in three classes of individuals, in terms of presence and intensity of pedophilic interest: teleiophilic, non-exclusively pedophilic, and exclusively pedophilic individuals.

structure in a large sample of sexual offenders who underwent phallometric assessment (Mackaronis, Byrne, Strassberg, Marcus, & Solari, 2011). However, their findings were ambiguous and did not support taxonic or dimensional structure.[2]

The discrepancy in findings may be partially explained by differences in the methods used and data conditions for taxometric analyses. The measures of pedophilic interest used by Schmidt and colleagues has an emerging body of empirical literature supporting their validity (Babchishin, Nunes, & Hermann, 2013; Banse, Schmidt, & Clarbour, 2010; Ó Ciardha, Attard-Johnson, & Bindemann, 2017; Schmidt, Babchishin, & Lehmann, 2017). In that study, the measures used were also within the range of data requirements for conducting taxometric analysis, suggesting the data conditions those authors were working with did little to skew the results. Stephens and colleagues (2017) used a mixture of validated measures of pedophilic interest along with measures with less well-established validity (i.e., the self-report measure used). In addition, one of the measures used by those authors was positively skewed beyond recommended limits, which can influence taxometric results (Waller & Meehl, 1998). A final potential explanation for Stephens et al.'s findings is that they used measures with victim age cut-offs that would have included men with hebephilic interest. The result is that they examined the latent structure of pedohebephilic interest, not pedophilic interest. Hebephilic interest may indeed have a dimensional latent structure, while pedophilic interest is taxonic (as suggested by Schmidt et al., 2013).

[2] Mackaronis, Strassberg, and Marcus (2011) conducted a taxometric analysis using subscales from the Multiphasic Sex Inventory-2 (MSI-2; Nicholas & Molinder, 1984), claiming to have assessed latent structure in pedophilic interest. However, the sexual obsessions and cognitive distortions subscales of the MSI-2 used in this study do not assess pedophilic interest. Given the choice of measure, we do not consider that study as having examined the latent structure of pedophilic interest.

A potential limitation to both studies, however, is the use of measures of sexual interest in children *relative* to sexual interest in adults. This method of assessing pedophilic interest is well-validated and widely used in research and clinical practice (McPhail et al., 2017). However, within the context of taxometric analysis, using relative measures of pedophilic interest may mix two distinct constructs within the analysis (i.e., pedophilic interest and teleiophilic interest). As a result, the use of relative measures of sexual interest may make it difficult to infer latent structure of sexual interest in children via taxometric analysis. Conceptualizing pedophilic interest as the sexual interest in children, without a comparison to sexual interest in adults, is consistent with relevant theory (e.g., Hanson, 2010; Seto, 2017) and may improve clarity of taxometric results.[3]

An additional explanation for the discrepant results is that the latent structure of pedophilic interest is more complex than the possibilities considered in these studies (i.e., dichotomous vs. dimensional). For instance, pedophilic interest may be trichotomous (i.e., consist of three ordered classes). Different classes of men may experience low, moderate, and high levels of pedophilic interest, with clear dividing boundaries between these three levels of sexual interest. Perhaps the clearest indication of a trichotomous structure in pedophilic interests available is the exclusive and non-exclusive subtypes (Cohen & Galynker, 2002; Finkelhor & Araji, 1986). Exclusively pedophilic sexual offenders have been found to be distinct in terms of rates of re-offending

[3] Importantly, taxometric analysis appears to answer the simpler question, "Is pedophilic interest dimensional or taxonic?" Most taxometric analyses we are familiar with pose this form of question (e.g., is psychopathy dimensional or taxonic?) and not a more complex form of the question, such as, "Is the bipolar construct, sexual interest in children relative to sexual interest in adults, dimensional or taxonic?" To continue with a psychopathy analogy, this more complex question would be analogous to asking, "Is the bipolar construct, psychopathy relative to being a saint, dimensional or taxonic?" Further to this, theoretical work has suggested that the strength of pedophilic interest should be separated from the exclusivity of pedophilic interest in order to understand the construct (Finkelhor & Araji, 1986). Diagnostically, strength of interest in children is the main consideration (APA, 2013); however, there is an evidence base suggesting relative interest in children is important for specific validity purposes (Blanchard et al., 2009). Based on these considerations, we have intentionally chosen to define pedophilia as the sexual interest in prepubescent children, without considering concomitant interest in adults. This will have ramifications for our methodology and the interpretation of the results.

against children (Biere, 1998) and exclusive pedophilic interest predicts sexual recidivism (Eher, Olver, Heurix, Schilling, & Rettenberger, 2015; Eher, Rettenberger, Matthes, & Schilling, 2010). In contrast, studies treating pedophilia dichotomously failed to establish predictive validity in this diagnostic approach (Eher et al., 2010; Kingston, Firestone, Moulden, & Bradford, 2007; Moulden, Firestone, Kingston, & Bradford, 2009; Wilson, Abracen, Looman, Pichea, & Ferguson, 2011). Within the context of taxometric studies, this hypothesis suggests that within a putative pedophilic taxon, there may be a second taxon distinguishable by high levels of arousal to children. Alternatively, this third class may only be distinguishable by interest in children relative to adults.

Trichotomous structure is important to consider in taxometric research because simulation studies suggest taxometric curves and fit indices can increase the number of ambiguous findings or provide conflicting results in the presence of a third taxon (McGrath, 2008; Walters, McGrath, & Knight, 2010). Visual examination of the taxometric curves in Schmidt and colleagues (A. Mokros, personal communication, March 1, 2017) suggests that the latent structure in their sample may be trichotomous, as opposed to dichotomous. That is, their results may suggest pedophilic interest consists of three ordered classes. The taxometric curves in Stephens and colleagues' analysis are more ambiguous. This is not necessarily surprising, as trichotomous structures are known to produce conflicting results across different datasets (McGrath, 2008).

2.1.2 Present Study

Currently, the results of taxometric analyses of pedophilic interest have been equivocal. Thus, the present study aims to extend the understanding of the latent structure of pedophilia in the following ways. First, the present study examines the latent structure of pedophilic interest using phallometric testing. Second, given the divergent findings in previous studies, we have

considered the hypothesis that pedophilic interest is neither dimensional nor dichotomous, but trichotomous. Third, given the absence of factor analytic studies examining whether female-oriented and male-oriented pedophilia load onto a single factor or separate factors, the present study examined the latent structure of pedophilic interest in three ways: both sexes combined, female-orientated, and male-orientated. Fourth, the present study conducted taxometric analyses in multiple datasets. Using multiple datasets may increase the confidence when latent structure is stable across samples. Last, using samples with different offence characteristics (e.g., nonsexual offenders, sexual offenders against adults, sexual offenders against children) and more theoretically and statistically justified phallometric indicators, the present study may resolve some of the validity issues present in past taxometric research using phallometric data (Mackaronis, Byrne et al., 2011).

2.2 Method

2.2.1 Samples

Precollected data from four samples of participants from three separate institutions in Canada were employed. Ethical approval for secondary analysis of these data was provided by the University of Saskatchewan Behavioural Research Ethics Board (certificate #Beh 16-167).

2.2.1.1 Institut Philippe-Pinel (IPP)

These participants are 632 sexual and nonsexual offenders who underwent phallometric testing at the Institut Philippe-Pinel de Montréal, in Montréal, Quebec, Canada, from 1984–2012. The sexual offenders are men who had committed sexual offences against adults, children, or both. The sample consisted of men who 1) were serving a custodial sentence and participated in treatment at the institution, 2) were under community supervision and followed on an outpatient basis, or 3) were assessed as part of presentence hearings. The sample underwent phallometric

assessments using a French translation of the child sexual violence auditory stimuli set (Barsetti, Earls, Lalumière, & Bélanger, 1998; Quinsey & Chaplin, 1988).

2.2.1.2 Regional Psychiatric Centre (RPC)

These participants are 261 federally incarcerated sexual offenders who underwent phallometric testing at the RPC in Saskatoon, Saskatchewan, Canada. The sexual offenders are men who had committed sexual offences against adults, children, or both. Phallometric testing was conducted for the purposes of risk and treatment need assessment. The phallometric test used by this lab employed slide-based stimuli (for description of procedure, see Canales, Olver, & Wong, 2009).

2.2.1.3 Regional Treatment Centre (RTC)

These participants are federally incarcerated sexual and nonsexual offenders who underwent phallometric testing at the RTC in Kingston, Ontario, Canada. The sexual offenders are men who had committed sexual offences against adults, children, or both. Phallometric testing was conducted for the purposes of risk and treatment need assessment. Three samples of offenders were assessed at this institution. The first sample consisted of 531 offenders who underwent phallometric assessment after 1993 (hereafter referred to as RTC 1). Within the RTC 1 sample, 382 offenders underwent phallometric assessment using audio-based stimuli (RTC 1: Audio) and 377 offenders underwent phallometric assessment using slide-based stimuli (RTC 1: Slide; Looman & Marshall, 2001). Within the RTC 1 dataset, 228 men received both audio and slide assessments. However, given the recommended sample size of 300 to perform taxometric analyses (Ruscio et al., 2006), we included these men in both the audio and slide datasets. The audio stimuli are the English version of the child sexual violence auditory stimuli set (Quinsey & Chaplin, 1988). Because a distinct slide-based stimuli set was used prior to 1993, 805 offenders

who underwent phallometric prior to 1993 were used as a standalone sample (RTC 2; Baxter, Marshall, Barbaree, Davidson, & Malcolm, 1984).

2.2.2 Phallometric Measures and Procedure

Phallometric tests measure changes in penile circumference while stimuli depicting different ages, sexes, and sexual activities are presented (Laws, 2009). Sexual interest in a certain age-sex category is indicated by increases in penile tumescence while attending to stimuli reflective of that age and sex. Much of the available literature shows that sexual offenders against children can be differentiated from other groups based on phallometric tests for pedophilic interest (e.g., Blanchard et al., 2001; Cantor & McPhail, 2015; McPhail et al., 2017). Phallometric testing is also a robust predictor of sexual recidivism by sexual offenders against children (Hanson & Morton-Bourgon, 2005; McPhail et al., 2017) and has been shown to have somewhat adequate reliability (test-retest r = .51, 95% confidence interval [CI] = .47, .55, k = 6, N = 1,265; McPhail & Olver, 2018).

The phallometric tests used either audio-based or slide-based stimuli. The audio-based stimuli involve an assessee listening to sexual interactions between two people, narrated in the first-person. The sexual interactions differ according to the age of the sexual partner (prepubescent child/adult), the sex of the partner (female/male), and the use of force involved in the sexual interaction (non-coerced/coerced) (Barsetti et al., 1998; Quinsey & Chaplin, 1988). In the IPP sample, there are four stimulus trials for which data are available: prepubescent female (non-coerced), prepubescent female (coerced), prepubescent male (non-coerced), and prepubescent male (coerced). In the RTC 1: Audio sample, there are 2 stimulus trials across 4 stimulus categories for which data are available: prepubescent female (non-coerced), prepubescent female (coerced), prepubescent male (non-coerced), and prepubescent male

(coerced)[4]. This resulted in the RTC 1: Audio having a total of 8 indicators that were used in the taxometric analyses. The slide-based stimuli involve projected photographs of a single nude or partially clothed person onto a screen in front of the assessee. The individuals in the slides vary by age, from early childhood to young adulthood, and sex (Baxter et al., 1984). For the slide-based phallometric tests, stimuli trials depicting female and male children aged 5, 8, and 10 were used in the present study. In the RTC 1: Slide, RTC 2, and RPC samples, there are 6 stimulus trials for which data are available for analysis.

Phallometric data for the samples were transformed into percent full erection scores. This data processing method involves dividing the maximum penile change during a stimulus trial by an estimate of full erection (i.e., 30mm of penile tumescence change; Becker, Stein, Kaplan, & Cunningham-Rathner, 1992b; Howes, 1995; Hunter & Goodwin, 1992) and multiplying the product by 100. Individuals who did not show arousal above 5 PFE during any stimulus trial were removed from the analyses.

2.2.3 Data Analyses

Three taxometric procedures were used to examine the latent structure of pedophilic interest. In each of the analyses, percent full erection scores during prepubescent child stimulus trials were used as the indicators of pedophilic interests.

2.2.3.1 Mean Above Minus Below a Cut (MAMBAC)

The MAMBAC procedure is based on the logic that if an indicator separates two latent taxa, then there is an optimal cutting score to distinguish the taxa (Meehl & Yonce, 1994; Ruscio et al., 2006). Using two indicators of pedophilic interest, this procedure uses multiple cutting

[4] We ran the taxometric analyses with and without the coerced stimuli removed from the audio stimuli datasets. The results did not change in a meaningful way and in order to retain a larger number of indicators, we report results including the coercive stimuli.

scores on one indicator to sort cases into two groups. Means scores on the second indicator for the group falling below the cut score are then subtracted from the mean scores for the group falling above the cut score. This subtraction is then repeated across many cut scores on the first indicator and these mean differences on the second indicator are plotted on a graph. Under ideal data conditions, taxonic latent structure results in peaked curves, while dimensional latent structure produces a concave curve. Trichotomous latent structure will result in twin peaked curves (McGrath, 2008).

2.2.3.2 Maximum Eigenvalue (MAXEIG)

The MAXEIG procedure requires three or more indicators of pedophilic interest. In this procedure, one variable serves as an input indicator, which is used to sort cases from lowest to highest according to score on the variable, and all other variables are used as output indicators (Meehl & Yonce, 1996; Ruscio et al., 2006; Waller & Meehl, 1998). In MAXEIG analyses, there are separate trials conducted in which each variable is used as the input indicator and the other variables are used as the output indicators. For example, if five variables are available, a MAXEIG analysis will include five trials in which each variable is selected to be the input variable while the other four variables serve as output variables. The input indicators are used to construct multiple subsamples of participants. In the present analyses, the sample was divided into multiple, overlapping subsamples of 100. In each subsample, the MAXEIG procedure computes an eigenvalue to determine the association between the 2 or more output indicators. This eigenvalue is derived from a variance-covariance matrix in which the diagonal is replaced with zeros, leaving only covariances in the matrix, and which represents the shared variance accounted for by the linear combination of indicators (Tabachnick & Fidell, 2013; Ruscio et al., 2006). The first and largest eigenvalue is plotted on a graph. In the presence of a latent taxon, a

subsample composed of all or mostly taxon or complement members will result in small eigenvalues because it is anticipated that variables will not covary within the taxon or complement class. Subsamples that are composed of a mixture of complement and taxon members will result in higher eigenvalues since taxon members tend to score high on the output indicators and complement members will tend to score low on the indicators output. Taxonic latent structure will produce peaked curves under ideal data conditions. Dimensional latent structure will produce eigenvalues across subsamples that are relatively consistent because subsamples are mixtures of individuals with varying levels of the trait. Trichotomous latent structure will result in twin peaked curves (McGrath, 2008).

2.2.3.3 Latent Mode Factor Analysis (L-Mode)

The L-Mode procedure involves conducting a factor analysis using multiple indicators of pedophilic interest that is constrained to a single factor solution (Ruscio et al., 2006; Waller & Meehl, 1998). The logic of this procedure is that scores on a factor will more validly separate taxon and complement members than scores on single indicators. The factor scores for each participant are plotted in a graph that has factor score on the *x*-axis and the relative frequency of cases at different factor scores on the *y*-axis. Taxonic latent structure will produce an L-Mode curve with a bimodal distribution (i.e., one peak represents complement member scores on the latent factor, the other peak represents taxon member scores) and dimensional latent structure will produce an L-Mode curve with a unimodal distribution. Trichotomous latent structure will result in a trimodal L-Mode curve (McGrath, 2008).

2.2.3.4 Comparison Curve Fit Index (CCFI)

While examining graphical output is a feature of taxometric analysis interpretation, the interpretability of taxometric curves is susceptible to the data conditions of the variables included

in the analysis. To ameliorate some of the difficulty in interpreting taxometric curves, a comparison curve fit index is computed. The CCFI is a fit index that measures the similarity between the research data being used in a taxometric analysis (i.e., phallometric scores) and boot-strap simulated dimensional or taxonic comparison data. This method simulates two sets of comparison data that have the same sample distributions, correlations between indicators, and indicator skew as the original research data (Ruscio et al., 2006). One set of comparison data are simulated assuming the underlying latent structure is taxonic, while the other set of comparison data assumes dimensional latent structure. These samples of simulated data undergo the same taxometric procedures as the original research data. The taxometric curves produced by the original research data are compared to the curves produced by the simulated dimensional data and simulated taxonic data, with a root mean square residual (RMSR) being computed for the distance of the research data curve from both simulated curves. A CCFI is computed during each taxometric analysis using the following formula: $RMSR_{dimensional}$ / ($RMSR_{dimensional}$ + $RMSR_{taxonic}$). The CCFI quantifies whether the research data is more similar to dimensional or taxonic latent structure, represented by the simulated data. CCFIs range from 0 to 1, with values from 0 to 0.399 indicating better fit for dimensional structure, values from 0.600 to 1 indicating better fit for taxonic structure, and values around .50 indicating ambiguous fit (Ruscio et al., 2006). All taxometric analyses were performed using the R taxometric program by J. Ruscio (2014).

2.2.3.5 Treatment of Phallometric Data

Phallometric data represent percent full erection, which quantifies, from 0-100, the maximum percent of full erection a man exhibited during each phallometric stimulus trial. Each stimulus trial was entered as an indicator into the taxometric analyses. For instance, there were

six different stimulus trials assessing pedophilic interest in the RTC 2 sample and the six trials were treated as separate indicators in the analyses. Taxometric analyses were run in three ways: 1) using all stimulus trials, 2) restricted to stimulus trials depicting female children, and 3) restricted to stimulus trials depicting male children. As there is currently little understanding whether phallometric stimuli load on separate age-sex factors or onto age factors, there appears to be little justification for not assessing the latent structure of the two sexes separately.

2.2.3.6 Analytic Plan

As the validity of variables used as indicators in taxometric analysis is a key consideration in interpretation of results, validity indices for the phallometric indicators are reported. In taxometric analysis, calculation of validity indices requires participants to be identified as putative taxon and complement class members. In order to establish putative taxon and complement class membership, the base rate classification technique was used because this method appears more accurate (Ruscio & Kaczetow, 2009; Ruscio, Ruscio, & Meron, 2007). Past taxometric studies of pedophilic interest have produced somewhat divergent taxon base rates (14.2% in Schmidt et al., 2013; 28.7% in Stephens et al., 2017), not to mention the differing conclusions regarding latent structure these studies resulted in. Other research estimates that the proportion of sex offenders classified as pedophilic is around 40% (Blanchard et al., 2001). This amount of ambiguity in the estimates of pedophilic base rates appears to have the potential to skew taxometric results. As a result, we chose not to use a specific base rate for a putative pedophilic taxon due to these divergent findings. Under these circumstances, the taxometric software program uses the taxon base rate estimated by the analyses to classify individuals into the putative taxon and complement class. The average estimated taxon base rate in the present

study was between 18.6% and 25.4% across datasets, which coheres closely with previous taxometric studies.

The validity indices reported here are skewness of phallometric indicators, the average correlation among phallometric indicators in the full sample, the average correlation among phallometric indicators in the putative taxon class and putative complement class (i.e., nuisance covariance), and the average standardized mean difference (i.e., Cohen's *d*) on the phallometric indicators between the putative taxon and complement class. As a general set of heuristics, for taxometric analyses to be optimally valid skew should be less than 1, correlations in the taxon and complement should be lower than $r = .30$, and between class validity should be greater than $d = 1.25$ (Ruscio et al., 2006). To aid interpretation of the results, when indicator validity estimates are not in these specified ranges, taxometric analyses result in higher rates of ambiguous findings (i.e., CCFIs between 0.40 and 0.60; Ruscio & Kaczetow, 2009). However, when indicator validity estimates are not within these specified ranges, there is a negligible increase in inaccurate results (i.e., a taxonic result when data are actually dimensional and vice versa; Ruscio & Kaczetow, 2009). These validity estimates will be reported in each dataset separately and the combined male and female stimuli analyses and the analyses for female stimuli and male stimuli separately.

Next, the CCFIs resulting from the MAMBAC, MAXEIG, and L-Mode analyses will be reported for each of the datasets separately. The CCFI values are reported for the combined male and female stimuli analyses and the analyses examining the female stimuli and male stimuli separately. The graphical output from the three taxometric procedures is also provided. Given the large number of analyses run across datasets, most of the graphical output is presented as

supplemental material and the interested reader is encouraged to visually inspect these taxometric graphs.

Because we considered latent structure beyond dimensional and dichotomous, we relied on simulation studies that provide taxometric graphical output for trichotomous structure to identify trichotomous structure in the graphical output in the present study (McGrath, 2008; Walter, McGrath, & Knight, 2010). Unfortunately, CCFIs do not provide an indication of whether trichotomous structure is present in the data. If trichotomous latent structure is indicated in the graphical output of the taxometric analyses, we pursued the possibility of trichotomous structure with post-hoc taxometric analyses using the procedure described in Ruscio and Ruscio (2004). To test for trichotomous structure, after the first set of taxometric analyses are run, those participants identified as belonging to the pedophilic taxon are used in a second set of taxometric analyses and the participants identified as belonging to the complement are removed from further analyses. This second set of analyses proceeds in a similar manner as the first: using taxon members, MAMBAC, MAXEIG, and L-Mode analyses are conducted. If these results indicate taxonic latent structure (i.e., graphical output look taxonic and CCFIs are greater than 0.60), this suggests the presence of a third class. If these results indicate ambiguous or dimensional latent structure, this suggests there is only the taxon and the complement identified in the first step of analysis (i.e., dichotomous structure). If dichotomous or trichotomous latent structure are indicated by these analyses, the identified taxa will be characterized using variables available in the datasets.

2.3 Results

Descriptive statistics and validity indices for the variables in the datasets are presented in Table 2-1 (at the end of Chapter 2). Given the elevated level of positive skew present in the data,

scores on all indicators above the 99th percentile were winsorized (Tabachnick & Fidell, 2013; Wilcox, 2005). Because elevated skew (i.e., skew > 1.00) and nuisance covariance (i.e., *r* in putative taxon or complement > .30) represent violations of taxometric assumptions, the following procedures were used to increase confidence and interpretability of the results: 1) a dual threshold for interpreting CCFIs (i.e., CCFIs less than .400 are interpreted as indicating dimensional structure, CCFIs greater than .600 are interpreted as indicating taxonic structure, and CCFIs between .400 and .600 are interpreted as ambiguous) and 2) imposing a multiple hurdles approach to interpreting taxometric results. A multiple hurdles approach involves interpreting taxometric results when the majority of CCFIs (i.e., 2 out of the 3 CCFIs produced by the MAMBAC, MAXEIG, and L-Mode procedures) or the mean of the three CCFIs exceed the dual threshold (Ahmed, 2010; Ruscio & Kaczetow, 2009; Ruscio et al., 2007). We recommend readers interpret the present findings using a conservative approach (i.e., using dual thresholds and a multiple methods approach).

CCFIs for the taxometric analyses of pedophilic interest, female-oriented pedophilic interest, and male-oriented pedophilic interest are presented in Table 2-2. The results were most consistent with taxonic latent structure. Pedophilic interest was indicated to be taxonic across 66.7% of the CCFIs (i.e., CCFI ≥ .60; 10/15 CCFIs), 20% were ambiguous (i.e., CCFI between .40 and .60; 3/15 CCFIs), and 13.3% supported dimensional structure (i.e., CCFI ≤ .40; 2/15 CCFIs). When using the majority method for identifying latent structure, 80% of the analyses indicated taxonic structure to pedophilic interest. For the averaged CCFIs, 60% supported taxonic structure and 40% were ambiguous. The weighted mean CCFI for the five datasets was .610 (*N* = 2,482). The taxometric curves for the RTC 1: Audio dataset are presented in Figure 2-1. We report the curves for this dataset because these illustrate the potential for trichotomous

structure most clearly. While a similar pattern is also detectable for other analyses, it is less clear (see A-1 to A-12 in Appendix A).

For female-oriented pedophilic interest, taxonic structure was indicated across 50% of the CCFIs (i.e., 6/12 CCFIs), while 41.6% were ambiguous (i.e., 5/12 CCFIs) and 8.3% supported dimensional structure (i.e., 1/12 CCFIs). When using the majority method for identifying latent structure, 50% of the datasets indicated taxonic structure to pedophilic interest. For the averaged CCFIs, 25% of the datasets supported taxonic structure and 75% were ambiguous. The weighted mean CCFI for the four datasets was .580 (N = 1,850). For male-oriented pedophilic interest, taxonic structure was indicated across 75% of the CCFIs (i.e., 9/12 CCFIs), while 8.3% were ambiguous (i.e., 1/12 CCFIs) and 16.6% supported dimensional structure (i.e., 2/12 CCFIs). When using the majority method for identifying latent structure, 100% of the datasets indicated taxonic structure to pedophilic interest. For the averaged CCFIs, 50% of the datasets supported taxonic structure and 50% were ambiguous. The weighted mean CCFI for the four datasets was .603 (N = 1,850). Notably, the IPP dataset did not contain enough phallometric trials to allow for analyses of sex-orientation.

2.3.1 Post-hoc Analysis

Visual inspection of the taxometric curves suggest trichotomous latent structure may also be present (see Figure 2-1; McGrath, 2008; Walters et al., 2010). Most strikingly, multiple L-Mode curves have trimodal distributions[5]. Following the procedure to assess for trichotomous latent structure outlined by Ruscio and Ruscio (2004), we conducted a second set of taxometric analyses using men classified as taxon members in the first round of analysis. To identify taxon

[5]As a secondary check, MAXEIG Bayesian classification probabilities were also produced and visually inspected. A number of the Bayesian classification probabilities indicated a subset of the samples had a moderate chance of being classified to the taxon, which is consistent with trichotomous latent structure (McGrath, 2008).

members in each sample, averaged taxon base rates produced by the three analyses were used, which were: IPP = 22.6%, RTC 1: Audio = 23.4%, RPC = 20.6%, RTC 1: Slide = 18.6%, RTC 2 = 25.4%. For example, in the IPP sample, the 22.6% men in the sample that showed the highest level of responding to child stimuli were classified as taxon members. Because sample sizes were not large enough to conduct the taxometric analyses in each dataset separately, the second step of analyses combined the RTC 2, RPC, and RTC1: Slides datasets (N = 332) and the RTC 1: Audio and IPP datasets (N = 228)[6].

Table 2-3 provides descriptive and validity estimates for the second step of taxometric analysis. The majority of validity estimates are within the expected ranges, suggesting skew and nuisance covariance was less problematic compared to the first step of analysis. CCFIs from the second step of analyses for pedophilic interest, female-oriented pedophilic interest, and male-oriented pedophilic interest are presented in Table 2-4. Pedophilic interest was indicated to be taxonic across 83.3% of the CCFIs (i.e., 5/6 CCFIs) while 16.7% of the CCFIs were ambiguous (i.e., 1/6 CCFIs) and 0% supported dimensional structure. When using the majority method for identifying latent structure, 100% of the analyses indicated taxonic structure to pedophilic interest; for the averaged CCFIs, 100% supported taxonic structure. The weighted mean CCFI for the five datasets was .610 (N = 2,482). For female-oriented pedophilic interest, taxonic structure was indicated across 100% of the CCFIs (i.e., 3/3 CCFIs) and the averaged CCFIs supported taxonic structure. For male-oriented pedophilic interest, taxonic structure was

[6] We combined these datasets given the similarity of the phallometric stimuli used. The RTC 1: Audio and IPP datasets are based on a phallometric procedure using the same auditory stimuli (Quinsey & Chaplin, 1988); however, the IPP stimuli were translated to French (Barsetti et al., 1998). The RPC, RTC 1: Slide, and RTC 2 datasets all contain phallometric data using slide stimuli that are similar in terms of the age rages of persons depicted in the slides. We additionally ran the second step of analyses using only the RTC 2 dataset (n = 204). These analyses were conducted to protect against differences in the results that may have been caused by combining datasets. Conducting taxometric analyses using taxon members in the RTC 2 dataset resulted in CCFIs ranging from .669 to .879, which is consistent with the findings reported in Table 2-4. All validity estimates for the RTC 2 dataset were within expected limits.

indicated across 100% of the CCFIs (i.e., 3/3 CCFIs) and the average CCFI supported taxonic structure. Notably, the combined audio datasets did not contain enough phallometric trials to allow for analyses of sex-orientation. Figure 2-2 presents the taxometric curves for the combined audio datasets. The curves from the audio datasets most clearly suggest taxonic structure. While the curves from the other datasets suggest a similar pattern, it is less clear (see Figures A-13 to A-20 in Appendix A). Taken together, the taxometric analyses indicate that the pedophilia taxon is itself taxonic, suggesting a trichotomous latent structure to pedophilic interest.[7]

In order to describe the three taxa, phallometric indices and demographic and offence characteristics were examined (see Table 2-5). It is important to note that these analyses proceeded in a post-hoc manner: while not every variable available in the datasets is analyzed here, variables that are descriptively important, of theoretical interest, available across datasets, or available for every individual in a dataset were pursued. When the same variable was available across multiple datasets, a weighted mean and pooled standard deviation were computed.

When compared to the other taxa, men in the second pedophilic taxon were younger, had a greater number of total and child victims, showed greater arousal to children of both sexes and female children by relative and absolute phallometric indices, and showed greater arousal to male children by an absolute phallometric index. A surprising result was that the second pedophilic taxon showed greater absolute arousal to adults when compared to the two other taxa (*d*s = 0.66 and 1.30) and the first pedophilic taxon showed somewhat higher arousal to adults

[7] Misclassifying complement members as taxon members in the second step of analysis inflates the risk of artificially identifying taxonic structure within the taxon identified in the first step of analysis (Ruscio & Ruscio, 2004). For this reason, in the second step of analysis, all members of the complement class should be excluded from analyses. To protect against artificially identifying taxonic structure within the taxon due to inclusion of complement class members, we reduced the taxon base rate because overestimating the taxon base rate risks complement members being falsely classified as taxon members. Running the second step of analysis with reduced taxon base rates did not change the CCFI results in a meaningful way and are not reported.

than the non-pedophilic taxon (d = 0.49). When compared with the non-pedophilic taxon, the first pedophilic taxon was younger and showed greater arousal to children by absolute and relative phallometric indices. To characterize the relative arousal patterns of the three taxa, the non-pedophilic taxon was teleiophilic, the first pedophilic taxon was non-preferentially pedophilic, and the second taxon was preferentially pedophilic. These descriptive terms will be used for the remainder of the article.

The increasing level of arousal to children and adults displayed by the two pedophilic taxa presents an interpretive puzzle and may be indicative of problems of sexual compulsivity/hypersexuality in the pedophilic taxa. Previous research has suggested a link between either sex drive, hypersexuality, sexual preoccupation, or sexual compulsivity and paraphilic interests generally (Bouchard, Dawson, & Lalumière, 2017; Cantor, Klein, Lykins, Rullo, Thaler, & Walling, 2013; Davis, 2017; Dyer & Olver, 2016; Kafka & Hennen, 2003; Långström & Hanson, 2006; Långström & Seto, 2006; Långström & Zucker, 2005; Sutton, Sratton, Pytyck, Kolla, & Cantor, 2015), and pedophilic interest specifically (Davis, 2017; Klein, Schmidt, Turner, & Briken, 2015). The RPC dataset contained a measure of sexual compulsivity (i.e., the Sexual Compulsivity item of the Violence Risk Scale-Sexual Offender version[8]; Olver, Wong, Nicholaichuk, & Gordon, 2007-2017), and sexual recidivism rates, which can serve as a proxy measure of hypersexuality. These variables, in tandem, provide further opportunity to examine the level of hypersexuality in the three taxa. There was a significant difference in sexual compulsivity among the three taxa (Mantel-Haenszel $\chi^2(1) = 17.83$, $p < .001$). The odds of sexual

[8] The Sexual Compulsivity item of the VRS-SO is a 4-point scale, with scores ranging from 0-3. We conducted group comparisons on the Sexual Compulsivity item in two ways. For the Mantel-Haenszel χ^2 test, the Sexual Compulsivity item score of 0–3 was used. In Table 2-5, the Sexual Compulsivity item was coded as present (scores of 2 and 3) or absent (scores of 0 and 1) and odds ratios were computed. On the VRS-SO, items rated as 2 or 3 are indicative of problem areas and are used in clinical practice to identify treatment targets for sexual offenders (Olver et al., 2007).

compulsivity problems in the preferentially pedophilic taxon was greater when compared to the teleiophilic (odds ratio [*OR*] = 13.34) and non-preferentially pedophilic taxa (*OR* = 4.89). The sexual recidivism rates did not differ between the non-pedophilic and non-preferentially pedophilic taxa (*OR* = 0.93); however, the odds of sexual compulsivity problems did differ between these two taxa (*OR* = 2.73). The preferentially pedophilic taxon had a sexual recidivism rate higher than the non-pedophilic (*OR* = 2.44) and non-preferentially pedophilic taxa (*OR* = 2.64). There was not a significant difference in sexual recidivism rate among the three taxa ($\chi^2(2) = 2.86$, $p = .24$, $\varphi = .11$); however, there was a significant difference in sexual recidivism rates in the preferentially pedophilic taxon compared to non-pedophilic taxon (Fisher's exact $p = .008$).

2.4 Discussion

The present research used phallometric test data from multiple samples of mixed sexual and non-sexual offenders to examine latent structure in pedophilic interest. The first step of taxometric analyses indicated that, across datasets, pedophilic interest was taxonic. Given the shape of the taxometric curves, we further considered that pedophilic interests may be trichotomous in structure, rather than dichotomous. In particular, the curves produced by L-Mode analyses suggested a third mode to the right of the larger complement-class mode, which is consistent with previous simulation studies examining the effects of trichotomous structure on taxometric curves (see Figure 6 in McGrath, 2008). When restricting taxometric analyses to putative taxon members, CCFI values and taxometric curve shapes supported taxonic structure. In particular, the L-Mode curves were more clearly bimodal, which is anticipated if two taxa are present in the data. These results indicate that the structure of pedophilic interest is trichotomous. This finding held when female-oriented and male-oriented pedophilic interest were examined separately.

To further understand the finding of a three-class structure in pedophilic interest, we conducted a series of post-hoc analyses to characterize the three taxa. The taxa were characterized by different levels of arousal to children, arousal to children relative to adults, and number of child victims. One main interpretation of these results is that the three taxa appear teleiophilic, non-preferentially pedophilic, and preferentially pedophilic. The preferentially pedophilic taxon also displayed higher arousal to adults compared to the other taxa. To us, this suggested the presence of problems with hypersexuality/sexual compulsivity in the exclusively pedophilic taxon. We found support for this post-hoc hypothesis.

Trichotomous structure in pedophilic interest has implications for interpreting the results of previous taxometric studies. The findings reported by Schmidt and colleagues (2013) support taxonic structure. However, the Bayesian classification probabilities and the L-Mode curve from that study (A. Mokros, personal communication, March 1, 2017) are relatively similar to the curves in the present research. The latent profile analysis conducted by those authors showed a two-class structure was the most parsimonious model; however, a three-class model provided fit statistics equivalent to the two-class model. These two findings may suggest that trichomotous latent structure may have been present in their data. However, latent profile analysis cannot determine whether the classes differ by degree or kind, so the suggestion of three ordered taxa from these analyses is not robust. Those investigators did not consider the trichotomous hypothesis, but they would not have had a large enough sample size to test for the presence of a second pedophilic taxon. Given these previous findings by Schmidt et al., a trichotomous result in our data is not completely unexpected.

The present findings are at odds with the dimensional result reported by Stephens and colleagues (2017). Those authors had a large sample and used a more diverse set of measures of

pedophilic interest than what was available in the present research. These aspects of their study speak strongly to the potential correctness of their result. One main explanation for the divergent results is that the measures used by Stephens and colleagues were relative measures that compared pedophilic interest to teleiophilic interest. At this point, it is unclear what effect considering pedophilic interest and teleiophilic interest simultaneously has on taxometric results. However, Stephens et al. did not appear to consider an alternative beyond dimensional or dichotomous latent structure, which may have limited their pursuit of novel findings present in those data. The shape of the curves presented in their study are relatively ambiguous and are not suggestive of trichotomous structure. Simulation studies suggest that CCFIs and taxometric curves become more ambiguous, or even misleading, when a third class is present in the data and provides a possible alternative explanation for the conflicting findings across studies (McGrath, 2008; Walters et al., 2010). Despite three taxometric studies conducted to date, there remains uncertainty which latent structure best characterizes pedophilic interest. This state of affairs demands further studies examining the latent structure of pedophilic interest.

Paraphilias have been conceptualized as involving three inter-related aspects: sexual self-regulation (e.g., ability to manage one's sexual behaviour), atypical sexual interests (e.g., interest in prepubescent children), and intensity of sexual interest (Hanson, 2010). To varying degrees, the present research provides support for this conceptualization. The main results indicate that the taxa were distinguishable in terms of the level of atypical sexual interest. In addition, the taxa displayed differing rates of sexual compulsivity and levels of overall arousal, suggesting that sexual self-regulation and intensity of sexuality were present to differing degrees in the taxa. To a lesser degree, the rate of prior offending increased across the taxa. Taken together, we found three taxa that were separable in terms of pedophilic interest and this distinction was associated

with intensity of sexual interests and sexual self-regulation problems. These findings cohere well with previous empirical literature suggesting an association between hypersexuality and paraphilic interests, including pedophilia (see Kafka (2010) for a review; Bouchard et al., 2017; Cantor et al., 2013; Davis, 2017; Dyer & Olver, 2016; Klein et al., 2015; Walton, Cantor, Bhullar, & Lykins, 2017). Given that the present research used phallometric measures of sexual interest, the finding that arousal to adult stimuli increased across the three taxa may be partially explained by the increasing rate of sexual compulsivity in the taxa. This, however, would not account for the differences in differing levels of relative interest in children found in the taxa.

Examining the potential interaction between these aspects of sexuality is an interesting and important avenue for future research. For instance, disentangling whether pedophilic interest is associated with high sex drive (e.g., frequency of sex acts, sexual preoccupation, impersonal sexual behaviour; Carvalho, Stulhofer, Vieira, & Jurin, 2015; Knight & Graham, 2017; Stulhofer, Jurin, & Briken, 2016), problematic sexuality (e.g., use sex to cope with negative affective states, compulsive sexual behaviour; Knight & Graham, 2017), or both may improve our understanding of this relationship and inform intervention efforts with these men.

When arousal to children relative to adults was examined in the three taxa, the two pedophilic taxa were non-preferential and preferential in their arousal to prepubescent children. The distinction between preferentially and non-preferentially pedophilic interest is an important consideration in clinical forensic practice. Research examining rates of sexual recidivism by pedophilic and non-pedophilic sexual offenders found that there is little difference between these two groups (Eher et al., 2010; Kingston et al., 2007; Moulden et al., 2009; Stephens et al., 2017; Wilson et al., 2011). In contrast, preferentiality/exclusivity of pedophilic interest has a strong relationship with sexual recidivism (Biere, 1998; Eher et al., 2010) and has been found to predict

sexual recidivism over and above well-validated measures of sexual recidivism risk (Eher et al., 2015). For instance, Eher and colleagues (2015) found exclusively pedophilic sexual offenders to have a sexual recidivism rate approximately five times that of non-exclusively pedophilic sexual offenders (21.2% vs. 3.9%). In a non-clinical sample of pedophilic men, exclusivity of pedophilic interest was associated with having committed a sexual offence (Bailey, Berhard, & Hsu, 2016).

In the present research, the two pedophilic taxa displayed different rates of sexual recidivism (50% vs. 27.5%) and the non-preferentially pedophilic taxon had a sexual recidivism rate equivalent to the teleiophilic taxon (27.5% vs. 29.1%). The elevated rate of sexual recidivism in the preferentially pedophilic taxon may also be partially explained by the finding that these men had co-occurring risk factors for acting on their sexual interest: high levels of pedophilic interest, preferential pedophilic interests, and sexual compulsivity. Future research may disentangle the influences of intensity of sexual interest, sexual compulsivity, and preferentiality of pedophilic interest when examining associations between pedophilic interest, sexual behaviour, and other constructs.

Further implications of a trichotomous structure in pedophilic interest remain to be examined. While theoretically interesting, research in multiple domains will help elaborate whether the two pedophilic taxa are meaningfully different in terms of developmental and neurobiological correlates, sexual behaviour, co-occurring mental health issues, and treatment response. For instance, sexual offenders in the two pedophilic taxa may be followed in longitudinal research to establish base rates of recidivism and the environmental, dynamic risk, and personality factors that are related with sexual recidivism and desistance in each taxon. For research evaluating the effectiveness of treatment, membership in one of the pedophilic taxa may

moderate treatment response and using a trichotomous classification may help identify those who respond differentially to treatment (Beauchaine, 2003).

2.4.1 Limitations

The samples and measures used in the current study limit our ability to generalize the results. All samples included were comprised of men who were involved in the criminal justice system, with the majority of men being incarcerated. Whether a trichotomous structure will be replicated in a sample more representative of the male population remains to be seen.

The present research relied on phallometric data, limiting our analyses to a single method of measuring pedophilic interest, whereas more conceptually distinct measures increase construct coverage (Ruscio & Ruscio, 2004). A separate issue that likely resulted from using conceptually similar indicators of pedophilic interest was the elevated level of nuisance covariance present in these data (we discuss this limitation in more detail below). While examining latent structure across multiple datasets improves confidence in the findings, trichotomous latent structure in pedophilic interest awaits replication, both with non-clinical samples and with measures other than, or in addition to, phallometric testing.

The data conditions in the samples were not ideal for taxometric analysis. The moderate level of positive skew (i.e., skew > 1.00) and elevated level of nuisance covariance (i.e., *r* in taxon or complement > .30) present in these data likely affected the findings. Simulation studies do allow us to identify, with some confidence, the effects. Under data conditions like those in the present study, curves from the taxometric procedures become increasingly distorted and ambiguous (Meehl, 1995b; Ruscio et al., 2006). Specifically, the peaks in taxonic MAMBAC and MAXEIG curves shift to the right when nuisance covariance in the taxon class is higher than in the complement class, while the taxon mode in L-Mode curves becomes more ambiguous

(Ahmed, 2010; Meehl & Yonce, 1994; Ruscio et al., 2006). MAXEIG appears to be the analysis most affected by high nuisance covariance (Ahmed, 2010), which appears to be true in our results. Positive skew imparts a similar effect on taxonic curves, flattening MAMBAC curve peaks and reducing differentiation of a right mode in L-Mode curves when a taxon is present in the data (Ahmed, 2010; Meehl & Yonce, 1994; Ruscio et al., 2006). The curves in Figure 2-1 and supplemental material tend to follow this pattern, with most taxonic peaks shifted to the right of the graphs. Under these conditions, taxometric curves derived from taxonic data are more likely to be judged as dimensional (Ahmed, 2010).

In data conditions similar to those found in the present study, CCFIs tend to be robust to violations of multiple statistical assumptions. For instance, CCFIs continue to be relatively accurate when skew raises above 1 (Ahmed, 2010; Ruscio et al., 2007). However, as nuisance covariance exceeds $r = .30$, the rate of ambiguous CCFIs increases, while the rate of inaccurate CCFI values (i.e., a CCFI value that indicates dimensional structure when the research data are taxonic and vice versa) remains low and relatively stable (Ruscio, Walters, Marcus, & Kaczetow, 2010). Figure 6 in Ruscio et al.'s (2010) simulation study shows the rate of correct, ambiguous, and incorrect CCFIs as the level of nuisance covariance in a dataset increases. Applying the rates found in Ruscio et al. (2010) to the level of nuisance covariance found in the present data, we would anticipate ~15–30% of the CCFIs to be ambiguous, ~2–18% to support dimensional structure, and ~62–90% to support taxonic structure, assuming taxonic structure is present in the data. In our first step of analysis, these are approximately the rates we found. In addition to this, the data conditions in the second step of analysis were generally improved and the CCFIs and curves were less ambiguous.

Some steps can be taken to limit the effects of skew and nuisance covariance on taxometric interpretation. A method to lessen the problems associated with skew (i.e., elevated Type I error rates) is to set a higher threshold for interpreting CCFIs (Ahmed, 2010; Ruscio et al., 2007; Ruscio et al., 2010). In the present study, we selected the most conservative threshold for interpreting CCFIs (i.e., CCFI > .600 indicate taxonic structure and CCFI < .400 indicated dimensional structure; relying on the majority rule and averaged CCFIs for interpretation) in order to strengthen the confidence in the findings under these data conditions.

2.4.2 Conclusions

Pedophilic interest was found to be trichotomous in structure. The implication of this finding is that a majority of men do not experience pedophilic interests, while a minority of men do. Within this group of men who experience pedophilic interest, our findings suggest some men are non-preferentially pedophilic and some men are preferentially pedophilic. The approach used in the present study is also instructive regarding interesting and important alternatives to consider when conducting taxometric analyses of human sexuality constructs. Considering trichotomous structure, or even more complex structures (see Borsboom et al., 2016), can provide a more accurate understanding of latent structure, and even has the potential to meaningfully alter the interpretation of taxometric results. Perhaps this lesson is summed up best by quoting Meehl (2004, p. 43), “No statistic is self-interpreting”. In the context of taxometric analysis, we suggest understanding latent structure goes beyond simple interpretation of CCFIs.

Table 2-1

Descriptive statistics and validity estimates for indicators

	Skew[a]	*r* in full sample[a]	*r* in taxon[a]	*r* in complement[a]	Validity[a]
Audio Stimuli Datasets					
IPP (*n* = 632)	1.96	.72	.43	.26	2.73
RTC 1 (*n* = 382)	1.57	.65	.30	.34	2.24
Female-oriented	2.07	.72	.44	.37	2.43
Male-oriented	2.14	.71	.38	.45	2.45
Slide Stimuli Datasets					
RPC (*n* = 261)	2.54	.58	.25	.29	2.28
Female-oriented	2.05	.81	.33	.49	4.25
Male-oriented	3.03	.72	.40	.32	3.26
RTC 1 (*n* = 402)	1.70	.55	.25	.28	1.87
Female-oriented	2.03	.66	.48	.23	2.45
Male-oriented	1.37	.56	.22	.20	2.06
RTC 2 (*n* = 805)	2.21	.58	.34	.27	1.82
Female-oriented	2.07	.67	.34	.28	2.42
Male-oriented	2.35	.60	.39	.19	2.22

[a]These values represent averages across the three taxometric procedures.

Table 2-2

Comparative curve fit indices across the full samples

	MAMBAC	MAXEIG	L-Mode	Majority	Mean CCFI
Audio Stimuli Datasets					
IPP (*n* = 632)	.763	.627	.681	3/3	.690
RTC 1 (*n* = 382)	.800	.416	.628	2/3	.614
Female-oriented	.716	.361	.625	2/3	.567
Male-oriented	.693	.380	.637	2/3	.570
Slide Stimuli Datasets					
RPC (*n* = 261)	.834	.622	.528	2/3	.661
Female-oriented	.672	.484	.427	1/3	.528
Male-oriented	.746	.727	.607	3/3	.693
RTC 1 (*n* = 402)	.709	.322	.609	2/3	.547
Female-oriented	.727	.751	.478	2/3	.652
Male-oriented	.690	.343	.600	2/3	.544
RTC 2 (*n* = 805)	.719	.400	.565	1/3	.561
Female-oriented	.671	.438	.598	1/3	.569
Male-oriented	.724	.474	.656	2/3	.618

Note. CCFI = Comparative Curve Fit Index; L-Mode = Latent Mode; MAMBAC = Mean Above Minus Below a Cut; MAXEIG = Maximum Eigenvalue.

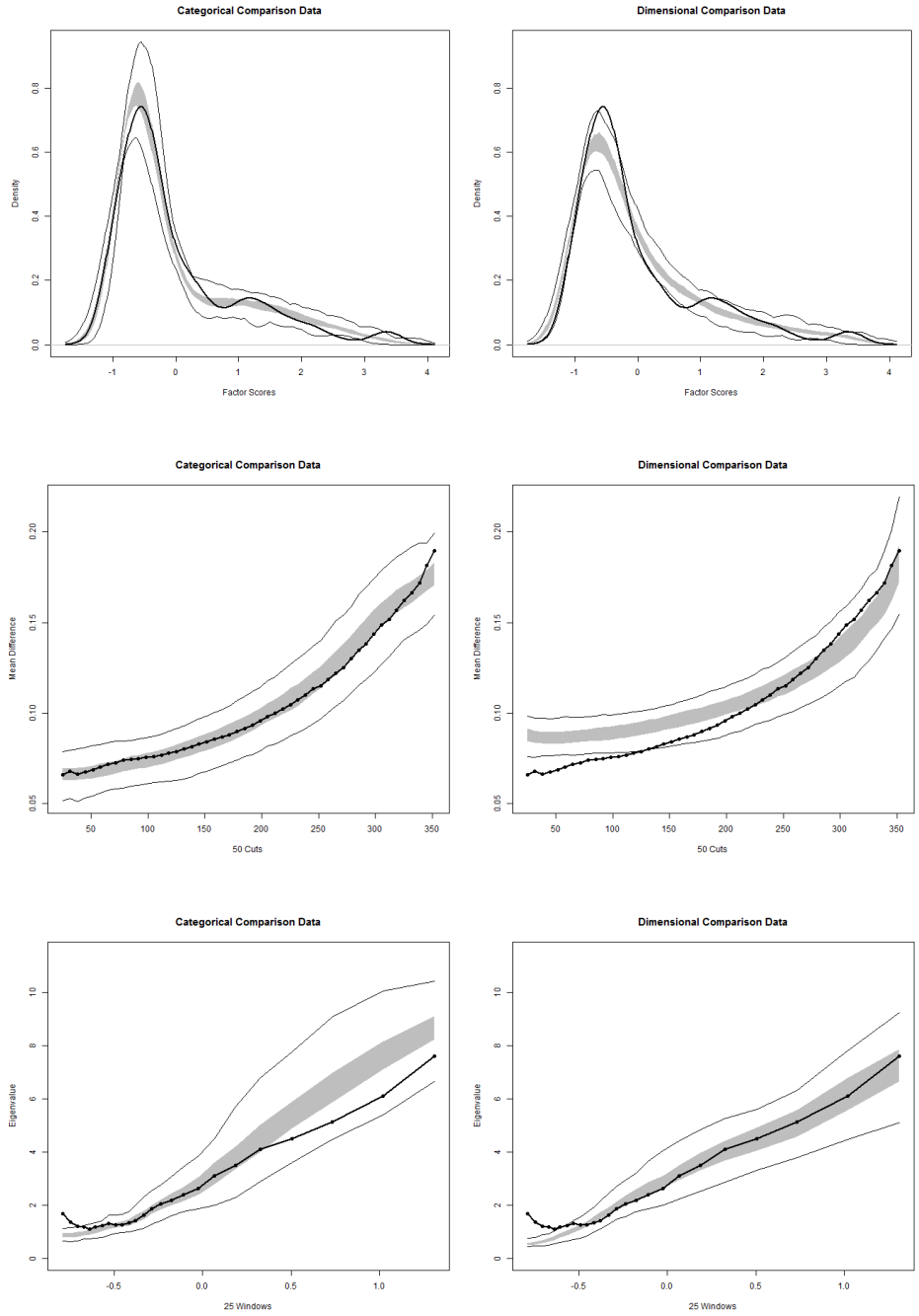

Figure 2-1. L-Mode, MAMBAC, and MAXEIG graphs for the RTC 1: Audio dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes.

Table 2-3

Descriptive statistics and validity estimates for indicators in the taxon

	Skew[a]	*r* in full sample[a]	*r* in taxon[a]	*r* in complement[a]	Validity[a]
Audio Stimuli Datasets					
Audio datasets combined (n = 228)	0.81	.70	.36	.31	2.23
Slide Stimuli Datasets					
Slide datasets combined (*n* = 332)	1.75	.38	.13	.08	1.85
Female-oriented	1.81	.69	.41	.32	2.92
Male-oriented	1.69	.56	.27	.15	2.42

[a]These values represent averages across the three taxometric procedures.

Table 2-4

Comparative curve fit indices in the pedophilia taxon

	MAMBAC	MAXEIG	L-Mode	Majority	Mean CCFI
Audio Stimuli Datasets					
Audio datasets combined (*n* = 228)	.753	.579	.663	2/3	.665
Slide Stimuli Datasets					
Slide datasets combined (*n* = 332)	.868	.765	.754	3/3	.796
Female-oriented	.797	.739	.692	3/3	.743
Male-oriented	.723	.618	.717	3/3	.686

Note. CCFI = Comparative Curve Fit Index; L-Mode = Latent Mode; MAMBAC = Mean Above Minus Below a Cut; MAXEIG = Maximum Eigenvalue.

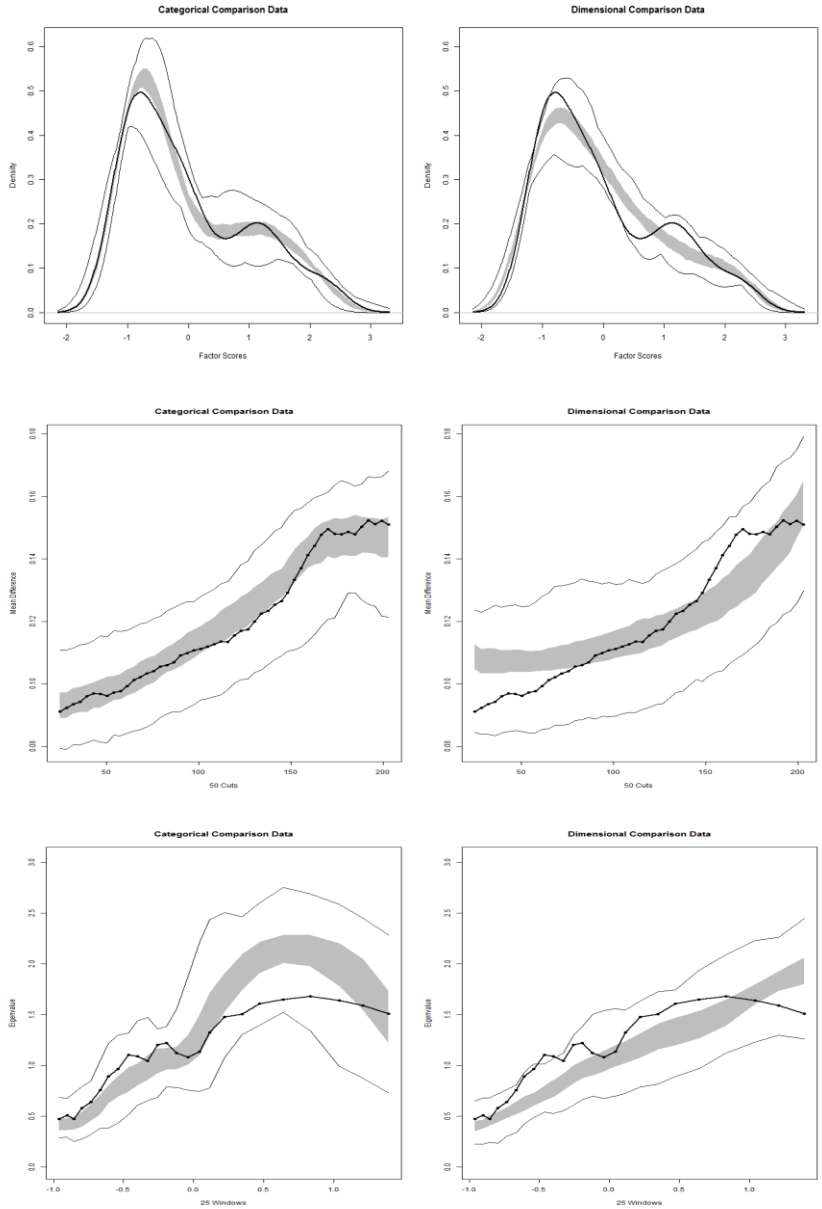

Figure 2-2. L-Mode, MAMBAC, and MAXEIG graphs for the combined audio dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes.

Table 2-5

Phallometric, demographic, and offence characteristics in the three taxa

	Non-pedophilic *M* (*SD*)	First Pedophilic *M* (*SD*)	Second Pedophilic *M* (*SD*)	*d* (95% CI) 3 vs. 1	*d* (95% CI) 3 vs. 2	*d* (95% CI) 2 vs. 1
Age	36.09 (9.91) *n* = 665	33.07 (9.55) *n* = 172	29.75 (8.57) *n* = 65	**–0.65** [–0.90, –0.39]	**–0.36** [–0.64, –0.07]	**–0.31** [–0.48, –0.14]
Pedophilia Index (PFE)	–6.74 (19.71) *n* = 1,824	2.35 (28.08) *n* = 388	11.81 (23.72) *n* = 157	**0.92** [0.76, 1.02]	**0.35** [0.16, 0.54]	**0.42** [0.31, 0.53]
Pedophilia Index (z-score)	–0.66 (1.58) *n* = 869	–0.17 (1.74) *n* = 195	0.67 (1.31) *n* = 69	**0.85** [0.60, 1.10]	**0.51** [0.23, 0.79]	**0.30** [0.15, 0.46]
Pedophilia Ratio (PFE)	0.77 (1.00) *n* = 1,822	1.20 (2.10) *n* = 388	1.77 (3.81) *n* = 156	**0.70** [0.53, 0.86]	**0.21** [0.02, 0.40]	**0.34** [0.23, 0.45]
Max child (PFE)	9.47 (7.73) *n* = 1,860	32.58 (23.99) *n* = 392	56.11 (25.77) *n* = 158	**4.51** [4.30, 4.72]	**0.84** [0.65, 1.03]	**2.14** [2.01, 2.26]
Max adult (PFE)	16.23 (20.91) *n* = 1,824	30.83 (27.00) *n* = 388	44.23 (27.38) *n* = 157	**1.30** [1.13, 1.47]	**0.49** [0.31, 0.68]	**0.66** [0.55, 0.77]
Max female child (PFE)	8.34 (7.10) *n* = 1,858	27.99 (20.46) *n* = 393	48.99 (24.48) *n* = 157	**4.22** [4.01, 4.42]	**0.97** [0.77, 1.16]	**1.84** [1.71, 1.96]
Female child index (PFE)	–8.31 (19.30) *n* = 1,822	–3.18 (24.01) *n* = 389	4.84 (23.83) *n* = 156	**0.67** [0.50, 0.83]	**0.33** [0.15, 0.52]	**0.25** [0.14, 0.36]
Max male child (PFE)	7.12 (4.99) *n* = 1,859	20.50 (17.84) *n* = 393	36.90 (25.02) *n* = 157	**3.52** [3.33, 3.72]	**0.81** [0.62, 1.00]	**1.53** [1.42, 1.65]
Male child index (PFE)	–9.07 (20.34) *n* = 1,822	–10.50 (29.24) *n* = 389	–7.11 (28.32) *n* = 156	0.09 [–0.07, 0.26]	0.12 [–0.07, 0.30]	–0.06 [–0.17, 0.04]
Number of victims	3.67 (4.75) *n* = 398	4.61 (8.62) *n* = 83	7.46 (13.46) *n* = 26	**0.67** [0.27, 1.07]	0.28 [–0.16, 0.73]	0.17 [–0.07, 0.40]
Number of adult victims	2.12 (5.38) *n* = 590	1.62 (3.68) *n* = 117	2.23 (5.18) *n* = 43	0.02 [–0.29, 0.33]	0.15 [–0.20, 0.50]	–0.10 [–0.30, 0.10]
Number of child victims	2.19 (6.40) *n* = 593	4.58 (9.27) *n* = 123	7.16 (13.74) *n* = 43	**0.70** [0.38, 1.01]	0.24 [–0.11, 0.59]	**0.34** [0.15, 0.54]
	% (*n*)	% (*n*)	% (*n*)	*OR* (95% CI)	*OR* (95% CI)	*OR* (95% CI)

				3 vs. 1	3 vs. 2	2 vs. 1
Sexual compulsivity problems	29.2% (47/161)	53.0% (18/34)	84.6% (11/13)	**13.34** [2.85, 62.51]	4.89 [0.94, 25.46]	**2.73** [1.28, 5.80]
Sexual recidivism	29.1% (57/196)	27.5% (11/39)	50.0% (7/14)	2.44 [0.82, 7.27]	2.64 [0.75, 9.26]	0.93 [0.43, 1.98]

Note. Bolded values indicate standardized mean differences that are statistically significant at $p < .05$. OR = Odds ratio. Odds ratios are computed for the categorical variables.

CHAPTER 3: FOLLOW-UP TAXOMETRIC ANALYSES TO FURTHER TEST LATENT STRUCTURE IN PEDOPHILIC INTEREST

3.1 Introduction

The taxometric study presented in Chapter 2 found initial support for pedophilic interest having a trichotomous latent structure. In the first set of taxometric analyses, a taxonic structure received the most support, suggesting there is a taxonic boundary that separates putative non-pedophilic from pedophilic men. To examine the latent structure within the pedophilic taxon, a second step of taxometric analyses were conducted which focused exclusively on men who were identified as pedophilic taxon members in the first step of analyses. This second step of taxometric analyses suggested a second taxonic boundary, which was interpreted as initial evidence of trichotomous latent structure. A limitation to the analytic approach taken in Chapter 2 is that it remains possible the first and second steps of taxometric analyses may have been identifying the same taxonic boundary. This would mean that the second step of taxometric analyses was merely reaffirming the taxonic boundary identified in the first step of analyses, not identifying a second taxonic boundary and trichotomous structure. The interpretive implication of this limitation is that pedophilic interest is indeed taxonic, but dichotomous.

This addendum chapter aims to further test the possibility that the latent structure of pedophilic interest is dichotomous, not trichotomous. The analyses conducted will take a similar approach to the second step of taxometric analyses as presented in Chapter 2. The main difference here is that the taxometric analyses will be conducted using samples with putative third taxon members removed. The remaining men in the samples can be considered to be members of the putative complement class and first pedophilic taxon, or those who were characterized as non-pedophilic and non-preferentially pedophilic in Chapter 2. The interpretive

possibilities are as follows: if this new set of analyses supports dimensional structure, this suggests that there is not a taxonic boundary separating men identified as non-pedophilic and non-preferentially pedophilic. Rather the distinction between these men is a matter of degree. This result would also suggest that the taxonic boundary that was identified in Chapter 2 separates preferentially pedophilic men from the non-preferentially pedophilic and non-pedophilic men. The taxon would then be composed of only preferentially pedophilic men, and pedophilic interest would have a dichotomous latent structure. However, if the new set of analyses provides support for taxonic structure, these results would provide further support for trichotomous latent structure, as this would reaffirm that the taxonic boundary indicated in first step of taxometric analyses in Chapter 2 was a boundary between non-pedophilic men and the combined non-preferentially and preferentially pedophilic men. Interpreting this result in conjunction with the second step of analyses in Chapter 2 would provide further initial validation that there are two taxonic boundaries in the latent structure of pedophilic interest (i.e., it has a trichotomous latent structure).

3.2 Method

3.2.1 Participants. The same five samples describe in Chapter 2 were included in the present analyses. The average base rate estimates, across the three taxometric procedures, in the audio datasets and slide datasets were used to identify men in the third taxon[9]. The men identified as belonging to the third taxon were removed from the samples, leaving only those who belonged to the taxa characterized as *non-pedophilic* and *non-preferentially pedophilic*

[9] For the audio datasets, the average of the three base rate estimates was .38 (MAMBAC = .37, MAXEIG = .35, L-Mode = .41). For the slide datasets, the average of the three base rate estimates was .16 (MAMBAC = .18, MAXEIG = .12, L-Mode = .19). The discrepancy between these two base rate estimates can likely be attributed to different base rates of third taxon members being present in the different samples. The average base rate for the audio datasets were then used to remove 50 men from the IPP sample and 24 men from the RTC 1: Audio sample. The average base rate for the slide datasets were then used to remove 15 men from the RPC sample, 18 men from the RTC 1: Slide sample, and 51 men from the RTC 2 sample.

(Institute Phillippe Pinel [IPP] *n* = 582; Regional Treatment Center [RTC] 1: Audio *n* = 349; RTC 1: Slide *n* = 314; RTC 2 *n* = 664; Regional Psychiatric Center [RPC] *n* = 247). The number of participants in the RPC sample falls below the suggested sample size for conducting taxometric analyses (i.e., *n* = 300). Although this sample was analyzed and the results are reported, generally, samples under 300 result in analyses that are less likely to be able to identify taxonic latent structure and more likely to produce ambiguous results (Beauchaine, 2007).

3.2.2 Analyses. Estimates of indicator validity for use in taxometric procedures will be provided. These estimates include magnitude of separation between putative complement and taxon members, indicator skew, and nuisance correlations (i.e., level of covariance between indicator scores within the putative complement class and taxon).

The three taxometric procedures described in Chapter 2 were conducted on each of the five samples (i.e., Mean Above Minus Mean Below a Cut [MAMBAC], Maximum Eigenvalue [MAXEIG], and Latent Mode Factor Analysis [L-Mode]). Comparative curve fit indices (CCFIs) are reported for the three taxometric procedures across the five datasets. A CCFI is an indicator of the fit between the research data and simulated taxonic and simulated dimensional data. These simulated comparison data are constructed using the same distributional properties as the research data (e.g., same level of indicator skew and nuisance covariance), with one set of simulated comparison data created under the assumption that the latent structure underlying the data is taxonic and one set of simulated comparison data is created under the assumption that the underlying latent structure is dimensional. In order to interpret whether the research data fits more closely to taxonic or dimensional simulated comparison data, CCFIs that are .600 or greater indicate the research data fits more closely to the taxonic simulated comparison data; whereas CCFIs that are .399 or less indicate the research data fits more closely to dimensional simulated

comparison data. In addition, the average of the CCFIs produced by the three taxometric procedures and the majority method (i.e., whether 2 or more out of the 3 CCFIs support one latent structure over the other) will be used to interpret these indices. The averaged CCFI values and whether a majority of individual CCFIs support one latent structure are typically the focus of interpretation.

3.3 Results

Estimates of indicator validity are reported in Table 3-1. Positive skew was somewhat elevated and nuisance covariance was slightly elevated in 2 datasets. The high level of separation between the complement and taxon in the datasets and the reasonable number of variables used are strengths within the datasets.

Table 3-2 provides the CCFI values for the taxometric procedures across the five samples[10]. In general, there was support for taxonic latent structure. For the averaged CCFIs and majority method, 80% of the results supported taxonic latent structure and 20% of the results were interpreted as ambiguous; there was no support for dimensional latent structure. Within the individual CCFIs, 80% supported taxonic structure, 20% were ambiguous, and none supported dimensional latent structure. The base rates estimates for the samples are provided in Table 3-3 and suggest that the identified taxon has a base rate between .20 and .24.

Examining the taxometric curves for the samples, a minority of curves aligned with prototypical shapes of curves produced by taxonic latent structure (see Figures B-1 to B-5 in Appendix B). The RTC 1 audio stimuli dataset once again produced taxometric curves that

[10] One phallometric indicator in the RTC 2 sample did not adequately differentiate between the putative complement and taxon members. Given this finding, the taxometric analyses were run with and without this indicator included, both results are provided in a note below Table 3.1 and in Table 3.2, respectively. Removing this indicator improved the clarity of the results, providing support for taxonic latent structure, whereas inclusion of this indicator resulted in more ambiguous findings. This is what we would anticipate when an indicator that does not adequately differentiate putative complement and taxon members is included in the analyses.

conform to prototypic taxonic curves (Beauchaine, 2007; Meehl & Yonce, 1994, 1996), such that the MAMBAC curve showed a small peak in the right-most side of the curve, the MAEXIG curve was exhibiting something resembling a peak, and the L-Mode curve was reasonably bimodal. The taxometric curves within the other datasets do not clearly resemble prototypical taxonic shapes, and this may be attributable to the presence of positive skew and nuisance covariance in the complement class.

3.4 Discussion

The results from this additional set of taxometric analyses provide support for taxonic latent structure when examining the complement class and members of a putative first pedophilic taxon. The question of whether this 'middle' taxon can be reliably identified was untested in Chapter 2, and the present results provide evidence of a taxonic boundary separating this middle taxon (i.e., non-preferentially pedophilic men) from complement members (i.e., non-pedophilic men). The results presented in Chapter 2 provide initial evidence that there is a second taxonic boundary that separates this 'middle' taxon from a third taxon (i.e., preferentially pedophilic men). The approach taken here and in Chapter 2, together, represent a more fulsome examination of the presence of two taxonic boundaries and provide support for the conjecture that pedophilic interest has trichotomous latent structure.

Table 3-1. Descriptive statistics and validity estimates for indicators

	Skew[a]	*r* in full sample	*r* in taxon[a]	*r* in complement[a]	Validity[a]
Audio Stimuli Datasets					
IPP (*n* = 582)	2.28	.49	.05	.29	2.32
RTC 1 (*n* = 349)	2.14	.58	.01	.39	2.19
Slide Stimuli Datasets					
RPC (*n* = 247)	3.30	.39	–.03	.31	2.16
RTC 1 (*n* = 314)	1.54	.39	–.06	.24	1.81
RTC 2[b] (*n* = 664)	1.92	.36	–.06	.25	1.79

[a]These values represent averages across the three taxometric procedures.
[b]The results in the table are from analyses when one variable was removed due to not differentiating between putative complement and taxon members (i.e., $d < 1.25$). When this variable was included in the analyses, the resulting validity estimates were: skew = 2.21, correlation in taxon = –.05, correlation in complement = .23, and validity = 1.67.

Table 3-2. Comparative curve fit indices across the full samples

	MAMBAC	MAXEIG	L-Mode	Majority	Mean CCFI
Audio Stimuli Datasets					
IPP (*n* = 582)	.783	.608	.744	3/3	.711
RTC 1 (*n* = 349)	.720	.613	.736	3/3	.690
Slide Stimuli Datasets					
RPC (*n* = 247)	.615	.505	.713	2/3	.611
RTC 1 (*n* = 314)	.685	.470	.426	1/3	.527
RTC 2[a] (*n* = 664)	.646	.630	.612	3/3	.629

Note. CCFI = Comparative Curve Fit Index; L-Mode = Latent Mode; MAMBAC = Mean Above Minus Below a Cut; MAXEIG = Maximum Eigenvalue.

[a]The results in the table are from analyses when one variable was removed due to not differentiating between putative complement and taxon members (i.e., $d < 1.25$). When this variable was included in the analyses, the three CCFIs were .627, .563, and .572; the average CCFI was .587.

Table 3-3. Base rate estimates across the datasets.

	MAMBAC	MAXEIG	L-Mode	Supplied BR	Mean BR within samples
Audio Stimuli Datasets					
IPP (*n* = 582)	.25	.15	.26	.16	.22
RTC 1 (*n* = 349)	.25	.17	.21	.18	.21
Slide Stimuli Datasets					
RPC (*n* = 247)	.22	.12	.26	.16	.20
RTC 1 (*n* = 314)	.24	.12	.34	.16	.23
RTC 2 (*n* = 664)	.36	.14	.21	.21	.24
Mean BR across samples	.26	.14	.26	.17	.22

CHAPTER 4

INTERVENTIONS FOR PEDOHEBEPHILIC INTERESTS IN SEXUAL OFFENDERS AGAINST CHILDREN: A META-ANALYTIC REVIEW

4.1 Introduction

The sexual abuse of children has wide-ranging adverse psychological, health, and financial impacts on victims and society (Cotter & Beaupré, 2014; Paolucci, Genuis, & Violato, 2001; Wang & Holton, 2007). Given the costs associated with child sexual abuse, understanding characteristics that increase individuals' likelihood of committing sexual offences against children and identifying effective treatments for these characteristics is a high priority.

Most theories of sexual offending include pedohebephilic interest as at least one risk factor that plausibly explains the initiation of sexual contact with children (for a review see Seto, 2018). Pedohebephilic interest connotes a sexual attraction in prepubescent or early pubescent children, characterized by sexual fantasies or sexual behaviour involving children or physiological arousal towards children (Cantor & McPhail, 2015). Pedophilic interest connotes a sexual attraction towards prepubescent children, who typically lack secondary sex characteristic development, while hebephilic interest connotes a sexual attraction to early pubescent children, who have begun to develop secondary sex characteristics; thus, pedohebephilia is an umbrella term that connotes both sexual attractions. Teleiophilic interest connotes a sexual attraction towards individuals who have reached sexual maturity and have developed secondary sex characteristics.

4.1.1 Assessment and Treatment Considerations of Pedohebephilic Interests

In forensic and criminal justice contexts, structured assessment approaches are employed with clientele charged or convicted for sexual offenses to assess their risk for future sexual violence, while intervention and management approaches are employed to mitigate that risk to

prevent further sexual violence (Bonta & Andrews, 2017). Although a range of clinical rating tools, diagnostic screens, and self-report measures exist to assess these individuals, a mainstay of assessing pedohebephilic interest has been phallometric testing, which assess changes in penile volume or tumescence during the presentation of erotic audiovisual stimuli. Changes in penile volume or tumescence during the presentation of erotic stimuli involving children is interpreted as indicating a sexual interest in children of a certain age and sex category; penile changes during the presentation of erotic stimuli involving consenting adults is interpreted as indication teleiophilic interest.

Although phallometric testing is not a risk assessment procedure per se, it assesses a clinically and forensically relevant psychosexological construct that can aid risk formulation and intervention. For instance, a recent meta-analysis of phallometric tests of pedohebephilic interests found that such interests are a strong predictor of sexual recidivism by men who have been convicted of sexual offences against children (McPhail et al., 2017). In turn, pedohebephilic interests fall under the rubric of what Mann, Hanson, and Thornton (2010), have termed psychologically meaningful risk factors; that is, biopsychosocial processes that are possible causes of sexual offending, predict maintenance of sexual offending, and processes that when treated and managed, lead to reduce sexual offending.

The risk-need-responsivity (RNR; Bonta and Andrews, 2017) model of effective correctional intervention explicitly links assessment and intervention through positing: 1) that the intensity of services (i.e., dosage) should be matched to the risk level of the client (risk principle), 2) that treatment should prioritize psychologically meaningful risk factors linked to criminal behaviour (need principle), and 3) service delivery should be tailored to the unique characteristics of clientele such as culture, motivation, and learning style among other areas

(responsivity principle). Pedohebephilic interests can be construed as an RNR-based construct, given that: 1) individuals assessed with a high level of such interests are higher risk for sexual offending and will require services of appropriate intensity, 2) pedohebephilic interest is a criminogenic need or psychologically meaningful risk factor associated with future sexual violence to be prioritized for intervention to reduce risk, and 3) sensitivity and discretion are required to promote client engagement and to encourage positive change in this domain.

Given the centrality of pedohebephilic interests in understanding sexual offending, several intervention approaches are used in sexual offense treatment programs (McGrath et al., 2010). For instance, behavioural interventions, based on operant conditioning principles, include aversion therapies in which a noxious stimulus is paired with arousal to children or reinforcement therapies in which adult stimuli are associated with a rewarding experience (e.g., masturbation; Marshall, O'Brien, & Marshall, 2009). In addition, pharmacological interventions, which include antiandrogen medications that reduce sex drive, can be provided as an adjunct or alternative to behavioural conditioning procedures (Garcia & Thibault, 2011). In research testing these interventions, phallometric testing may be used not only to assess the presence of pedohebephilic interests, but also as a treatment tool to monitor changes in the capacity to interrupt or inhibit arousal.

4.1.2 Present Study

The present study is a meta-analytic review of the effectiveness of interventions for pedohebephilic arousal. We examine treatment effects for intervention types across developmental stage of sexual offending persons, sexual offense subgroup, and ages of individuals depicted in erotic stimuli. Given the status of phallometric testing as an established

assessment measure of pedohebephilic interest and that it is used in much of the treatment literature, the present meta-analytic review will focus on studies using phallometric testing.

4.2 Method

4.2.1 Study Inclusion

Studies were included in the present meta-analysis if the research: 1) Included a sample of adult or adolescent sexual offenders against children, defined as men who had committed a sexual offence against a child under 15 years of age (Cantor & McPhail, 2015). Samples that had offended against related children, unrelated children, female children, or male children were included. 2) Reported data from a phallometric assessment of pedohebephilic interests. 3) Provided an intervention targeting pedohebephilic interests in a sample of sexual offenders against children. 4) Included sufficient statistical information to calculate a within-subjects, pre- to post-treatment change effect size (ES) statistic in a treatment group.

4.2.2 Literature Search

A literature search to identify eligible studies was conducted through March of 2017. We systematically searched multiple databases, including Pro-Quest Dissertations and Theses Global, PsycINFO, Web of Science, and PubMed. Searches included combinations of the following terms: “phallometry”, “penile plethysmography”, “PPG”, “sexual arousal”, “deviant arousal”, “deviant sexual interest”, “sexual preference”, “sexual offender against children”, “child molester”, “child sex offender”, “pedophile”, “pedophilia”, and “treatment”. A search of governmental agency websites; journal table of contents; conference programs; and reference lists from relevant review articles, books, and book chapters was also conducted.

4.2.3 Data Extraction and Coding

4.2.3.1 Study and sample level characteristics. Studies were coded for publication status, year of publication, setting of in which treatment was provided (prison/institution, community, or combination of both), and country.

4.2.3.2 Study design-level characteristics. Studies were coded according to the type of research design used, which included: single case designs, non-randomized single treatment group design, non-randomized treatment and control group design, and randomized control trial. Single case designs were included in order to capture a relatively large literature assessing the effect of behavioural treatments on pedohebephilic interests (k = 18). Data from follow-up assessments was also coded.

4.2.3.3 Treatment-level characteristics. Interventions were coded as behavioural, cognitive-behavioural, pharmaceutical, comprehensive treatment programs, eye-movement desensitization and reprocessing, or a combination of interventions. Specific techniques within these intervention types was also coded. Behavioural treatments included masturbatory reconditioning, olfactory reconditioning/aversion, covert or vicarious sensitization, and satiation. Cognitive-behavioural interventions coded for included thought-stopping, self-talk, identify automatic thoughts related to sexual arousal, and cognitive restructuring. Pharmaceutical interventions coded for included medroxyprogesterone acetate, cyproterone acetate, luteinizing hormone-releasing hormone analogues, and antipsychotics or selective serotonin reuptake inhibitors.

4.2.3.4 Phallometric assessment characteristics. Phallometric tests for pedohebephilic interest differentiate sexual offenders against children from other groups based (Cantor & McPhail, 2015; McPhail et al., 2017), are robust predictors of sexual recidivism, and have somewhat adequate reliability (see below). In the present study, the age of persons depicted in

stimuli used during phallometric testing was coded. Two age categories were used to identify child stimuli: prepubescent stimuli (aged 10 years and younger) were coded as *pedophilic* and pubescent stimuli (aged 11–14 years) were coded as *hebephilic*. These age ranges were adopted because these age ranges reflect when pubertal changes occur for children, which is typically before or around age 11 (see Cantor & McPhail, 2015). Some studies describe child stimuli without providing specific ages for the children, complicating the classification of the age of the subjects depicted in the stimuli. In these cases, the stimuli were coded as *pedohebephilic*. In addition, the *pedohebephilic* category includes stimuli coded as pedophilic and hebephilic and groups stimuli depicting individuals under the age of 15 into a superordinate category. Stimuli depicting adults were coded as *teleiophilic*. Phallometric stimuli were also coded according to the type of phallometric data reported (raw changes in penile circumference, percent full erection data, z-score data, indices derived from percent full erection data, or indices derived from *z*-score data). Data type was used to inform the normative comparison analysis (see below).

4.2.3.5 Moderating variables. Recent taxometric research has found pedophilic interest to be taxonic and dimensional (McPhail, Olver, Brouillette-Alarie, & Looman, 2018; Stephens, Leroux, Skilling, Cantor, & Seto, 2017; Schmidt, Mokros, & Banse, 2013). The main meta-analyses conducted model pedohebephilic interest as a dimension. However, a recent taxometric analysis has found a three-taxa structure to pedophilic interest, with the taxa being characterized as teleiophilic, non-preferentially pedophilic, and preferentially pedophilic (McPhail et al., 2018). Because latent structure can have implications for the course of treatment and how treatment effectiveness is assessed, we examined whether taxon membership moderates treatment effects. To assess whether taxon membership moderates treatment effects, the highest pre-treatment percent full erection score (PFE) in a sample was used to categorized studies into

one of three taxa, using PFE cut-scores differentiating the taxa provided by McPhail et al. (2018). The criterion for categorizing samples into taxon 1 was the highest mean pretreatment PFE score to a child stimulus trial being below 11.6. For taxon 2, these samples' highest mean pretreatment PFE score was between 11.6 and 24.8; for taxon 3, the criterion was the highest pretreatment PFE score greater than 24.8. These PFE cutscores represent the weighted averaged PFE scores, across the five samples included in McPhail et al. (2018), the separated the samples into the three taxa, using the taxa base rates produced by the taxometric procedures. Taxon membership was used as a moderator variable of treatment effects for all treatments combined, behavioural and comprehensive treatments combined, and behavioural treatments across pedohebephilic, pedophilic, and teleiophilic interests.

4.2.3.6 Risk of bias. Risk of bias in primary studies was assessed using the ROBINS-I tool, which provides a means to systematically assess risk of bias in non-randomized studies of interventions (Sterne et al., 2016a). Raters relied on descriptions of the items in the ROBINS-I guidance manual (Sterne et al., 2016b). Six of the seven domains of bias on the ROBINS-I were coded because the *selection bias* domain was not relevant to the included studies. Each domain is rated as having a low, moderate, serious, or critical risk of bias. The coder is asked to make two global ratings: the overall risk of bias present in the study and the direction of the bias. All single-group, pre-post design and single-case design studies were rated as having a "critical" risk of bias and that the direction of the bias was unpredictable.[11]

4.2.3.7 Interrater reliability. Interrater reliability analyses were based on ten studies, chosen at random, coded by two independent raters. The coders were aware of which studies

[11] It was our intent in using the ROBINS-I was to conduct sensitivity analysis using the risk of bias assessment. However, the large majority of studies were rated as being at critical risk of bias, making sensitivity analysis uninformative.

were being used for interrater reliability analysis. Interrater reliability ranged from κ = .53 to 1.00 for categorical variables and intraclass correlation values ranged from .82 to 1.00 for continuous variables.

4.2.4 Analytic Approach

4.2.4.1 Calculating effect sizes. Pre- and posttreatment means and standard deviations, *p*-values for pre- to posttreatment change, *t*-values for pre- to posttreatment change, and differences in means were coded from the studies in order to calculate ESs. A sizable minority of studies presented average phallometric data in figures and in order to include these studies, we used WebPlotDigitizer Version 3.8 (Rohatgi, 2018) to extract data from figures. The use of this program in meta-analytic reviews is recommended as a method to capture more study data (Burda, O'Connor, Webber, Redmond, & Perdue, 2017). Data from participants across single case design studies were averaged and a standard deviation for all participants in these studies was computed. This resulted in the single case design studies producing a single ES estimate for behavioural interventions.

A common issue in intervention studies is that many do not report a pre- to posttreatment correlation, which are needed to compute a within-subjects ES. Some authors recommend imputing a pre- to posttreatment correlation of r = .70 (Rosenthal, 1993); however, this approach can introduce bias (Cuijpers, Weitz, Cristea, & Twisk, 2017). In order to compute more accurate ESs for this meta-analysis, we reviewed studies included in McPhail et al. (2017) and additional research to identify studies that report test-retest correlations for phallometric tests for pedohebephilic and teleiophilic interests. We found six studies reporting test-retest correlations and conducted a meta-analysis of these data (see Appendix C.2 for full citations). For phallometric tests for pedohebephilic interests, the aggregate test-retest correlation was r = 0.51

(95% confidence interval [CI] = 0.47, 0.55), $Q = 4.80$, $I^2 = 0.00$, $k = 6$, $N = 1{,}256$; for teleiophilic interests, the aggregate test-retest correlation was $r = 0.43$ (95% CI = 0.26, 0.57), $Q = 2.89$, $I^2 = 0.00$, $k = 5$, $N = 124$. We imputed one of these two values for computing ESs, depending on the age of persons depicted in the phallometric stimuli.

4.2.4.2 Aggregating effect sizes. Data from the include studies were aggregated using both fixed-effect and random-effects meta-analysis. Fixed-effect meta-analytic results are conceptually restricted to the particular set of studies included in the meta-analysis, while random-effects meta-analytic results allow for more confidence in generalizing to the population the current sample of studies is drawn from. When variability across studies is low (i.e., $Q <$ degrees of freedom), random-effects and fixed-effect meta-analysis produce identical results. When the analysis includes a small number of studies ($k < 30$), greater interpretive weight should be given to fixed-effect rather than random-effects analyses because the between-study variability estimate necessary for random-effects analyses loses precision (Schulze, 2007).

Multiple studies reported data on two or more phallometric outcomes. Two approaches were used for studies with multiple outcomes. In the first approach, ESs from multiple outcomes were averaged within studies; this method is used in most of the analyses reported below. For example, Bradford and Pawlak (1993) reported pre- and posttreatment means for three phallometric stimuli and the mean ES of these three outcomes was used in this first method. In the second approach, we selected the phallometric outcome for which the study sample showed the highest average arousal at pretreatment. For example, the sample in Bradford and Pawlak (1993) showed the highest pretreatment arousal to stimuli depicting sexual activity with a passive child and this was the ES selected in the second approach.[12]

[12] We included this second method of dealing with multiple outcomes within studies because sexual offenders against children are not expected to show high levels of arousal to all phallometric stimuli depicting sexual activity

When two or more ESs were available, a meta-analytic aggregate effect was computed. Conducting a meta-analysis using only two ESs may produce an inaccurate estimate of dispersion and CIs; for this reason, when an aggregate effect is based on only two ESs, these should be interpreted with caution and as preliminary estimates (Borenstein, Hedges, Higgins, & Rothstein, 2009). Estimates of the heterogeneity from the fixed-effects model are reported. The *Q* statistic indicates whether the observed heterogeneity among individual ESs is statistically significant. I^2 indicates the proportion of the observed variability between studies that is due to factors beyond spurious variation (Borenstein, et al. 2009), and this statistic can be interpreted as an estimate of the amount of inconsistency in the findings of the studies included in a meta-analysis (Higgins, Thompson, Deeks, & Altman, 2003). Outlier analyses were conducted using the following criteria: four or more studies contributed to the mean ES, the *Q* statistic was significant ($p < .05$), an individual study's ES was the most extreme value, and an individual study's ES accounted for 50% or more of the *Q* value (Whitaker et al., 2008). Meta-analytic analyses were conducted using Comprehensive Meta-Analysis, Version 3.0 (Biostat, 2014).

Planned moderator analyses were conducted using taxa membership to group samples. In moderator analyses, the *Q* statistic is portioned into $Q_{between}$ and Q_{within}. $Q_{between}$ reflects the variability explained by the moderator variable (between-level variability) and Q_{within} reflects the pooled within-level variability (unexplained variability) (Borenstein et al., 2009). $Q_{between}$ follows a χ^2 distribution with $x - 1$ degrees of freedom, where x is the number of levels in the moderator.

4.2.4.3 Publication bias. Analyses for publication bias was assessed using the trim-and-fill method (Duval & Tweedie, 2000), Egger's test of the intercept (Egger, Smith, Schneider, &

with male and female children. Including in a meta-analysis effect sizes for phallometric stimuli to which the sample did not show high levels of responding may artificially reduce the effect of treatment, as treatment would not be expected to decrease arousal to stimuli that a person did not find arousing in the first place. However, examining the highest response can also artificially introduce artifacts related to regression to the mean.

Minder, 1997), and visual inspection of the funnel plot. The trim-and-fill method assesses whether studies with negative results are missing from a meta-analysis and can provide an estimate adjusted for missing studies. Egger's test uses a regression model to detect publication bias; if bias is present, the intercept in the model will deviate from zero. We conducted these analyses when appropriate to do so ($I^2 < 50$, non-significant Q, 5+ studies, at least one effect is significant; Ioannidis & Trikalinos, 2007) and report results when publication bias was present.

4.2.4.4 Benchmarking. Most studies used single group, pre-posttreatment designs. To ameliorate some of the limitations in using such studies in meta-analyses, we constructed natural history benchmarks for pedohebephilic and teleiophilic interests against which to compare treatment effects. These benchmarks provide a means of determining if treatment effects are greater than change due to natural history processes (Minami, Serlin, Wampold, Kircher, & Brown, 2008). Benchmarks for pedohebephilic interests were created by aggregating pre-post phallometric scores in three waitlist control groups and two test-retest samples (see Appendix C.3 for full citations). A meta-analysis of these five samples resulted in a pedohebephilic interests natural history benchmark of $g = 0.113$ (95% CI = –0.044, 0.269, $Q = 6.23$, $I^2 = 35.80$, $k = 5$, $N = 152$). Using two samples, a teleiophilic natural history benchmark was also created ($g = –0.076$, 95% CI = –0.309, 0.158, $Q = 3.09$, $I^2 = 67.64$, $k = 2$, $N = 74$). A range-null test was used to test whether treatment effects were beyond a critical value and can be interpreted as being statistically significantly greater than natural remission (Minami et al., 2008). This range-null test follows a non-central t distribution with $N – 1$ degrees of freedom. Non-centrality parameters, t-critical values, and g-critical values were derived using formulas presented in Minami et al. (2008). A predetermined margin for a clinically trivial difference was selected ($g =$

0.20) and aggregate treatment effects were considered clinically relevant if the ES was at least one-fifth of a standard deviation larger than the natural history benchmark.

4.2.4.5 End-state normative comparisons. Benchmarking allows for some confidence in identifying treatment effects that are greater than natural remission. However, these analyses do not address how well sexual offenders against children are functioning at the end of treatment. To address this limitation, end-state normative comparisons were conducted (Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999; McAleavey et al., 2017; McEvoy & Nathan, 2007). Normative comparison data were constructed by aggregating data from samples of men without a history of sexual offending reported in McPhail et al. (2017). Normative comparison data, in the form of means and standard deviations, were constructed for pedohebephilic, pedophilic, and teleiophilic interests (see Appendix C.4 for full citations).[13] Posttreatment means and standard deviations from treatment studies were used to construct weighted means and pooled standard deviations across the three sexual interests.[14]

The equivalence of means in the normative and treatment samples was tested using the two one-sided test procedure (Kendall et al., 1999; Lakens, 2017; van Wieringen & Cribbie, 2014). This method requires selecting a range of closeness (δ) that specifies the range within which group differences must fall in order to be considered equivalent. In the context of phallometric tests, there is little guidance available for what an appropriate interval would be to make the

[13] Phallometric data are not reported in a standardized format across studies. However, procedures for computing different data types are well-established and each type is calculated similarly across phallometric studies. Given this variation, we constructed normative means and standard deviations for percent full erection (PFE), z-scores, indices, and z-score indices data types. This was done so that we could match treatment study data to a normative comparison using that same phallometric data type. For example, a number of treatment studies reported posttreatment scores as PFE, and we compared the weighted mean and pooled standard deviation of these samples to the normative comparison studies reporting sample data as PFE.

[14] Seventeen treatment studies reported means and standard deviations for post-treatment phallometric data in treatment samples. The other studies did not report both means and standard deviations and could not be included in these normative comparison analyses.

normative comparisons. In this situation, multiple δ are selected and used in equivalence tests (van Wieringen & Cribbie, 2014). On reviewing the standard deviations in the normative samples ($SD_{normative\ sample}$) and the results in McPhail et al. (2017), we selected two intervals, $\delta = 0.5SD_{normative\ sample}$ and $\delta = 0.75SD_{normative\ sample}$, for the equivalence tests. The results for the 0.5*SD* and 0.75*SD* were identical and these results are presented together.

4.3 Results

4.3.1 Study Characteristics

Twenty-three studies describing treatment effects for samples of sexual offenders against children, producing 197 unique ESs (*MD* = 4) and including 1,045 sexual offenders against children (*MD* = 25), were included in the analyses. Of these 23 studies, 74% were published (*k* = 17), 48% provided treatment in an in-patient setting (*k* = 11) and 44% provided treatment in an outpatient setting (*k* = 10), and 74% used a non-randomized single treatment group design (*k* = 17), while 2 studies used random assignment in the design. Eighteen studies reporting single case designs, including 26 sexual offenders against children, were include in the single case designs. Of these 18 studies, 94% were published (*k* = 17), 67% provided treatment in an in-patient setting (*k* = 12), and 22% provided treatment in an outpatient setting (*k* = 4). See Table 4-1 for more detailed information from each study included in the analyses (see Appendix C.1 for full citations).

When interpreting the direction of ESs, a positive ES for pedohebephilic or pedophilic interests indicates that the treatment group showed lower levels of arousal to child stimuli from pre- to posttreatment. For teleiophilic interests, a positive ES indicates the treatment group showed higher levels of arousal to adult stimuli from pre- to posttreatment.

4.3.2 Overall Effect of Interventions

4.3.2.1 Behavioural treatments. The meta-analytic results for the effect of intervention on pedohebephilic, pedophilic, and teleiophilic interests are shown in Table 4-2. There was a positive effect for behavioural treatments for pedohebephilic and pedophilic interests; this effect was increased with the inclusion of single case design studies. When results were restricted to ESs derived from the samples' highest response to stimuli depicting children, the positive effect of behavioural treatments was large (g = 0.79 [0.63, 0.96], Q = 15.5, I^2 = 28.8, k = 12, N = 183). Behavioural treatments had little effect on increasing phallometric responding to adults. The treatment effect on pedohebephilic and pedophilic interests was clinically and significantly greater than the natural history benchmark critical value (g_{cv} ranged from = 0.43 to 0.66; all p < .01; see Appendix D, Table D-1). The treatment effect for pedohebephilic interest was also greater than the natural history benchmark using the samples' highest response to stimuli depicting children (g_{cv} = 0.50, p < .01, see Appendix D, Table D-2).

Two studies also examined the effect of behavioural treatments from pretreatment to follow-up and from posttreatment to follow-up. The magnitude of change from pretreatment to follow-up was g = 0.74 [0.40, 1.08], Q = 2.8, I^2 = 64.5, k = 2, N = 39). In addition, treatment gains were maintained from posttreatment to follow-up (g = 0.12, 95% CI [–0.18, 0.43], Q = 3.4, I^2 = 70.7, k = 2, N = 39).

4.3.2.2 Pharmacological treatments. Pharmacological treatments showed a similar positive effect on pedohebephilic interests (Table 4-2); however, there were too few ESs available to evaluate these treatments' effect on pedophilic and teleiophilic interests. Pharmacological treatment effects were greater than the natural history benchmark critical value (g_{cv} = 0.63; p < .05; Appendix D, Table D-1). A similar result was found when the treatment effects were examined using the samples' highest response to stimuli depicting children (g =

0.70 [0.20, 1.30], Q = 2.1, I^2 = 0, k = 4, N = 32; benchmark g_{cv} = 0.64, p < .05, Appendix D, Table D-2).

4.3.2.3 Comprehensive treatments. Comprehensive treatment programs showed a small, yet significant positive effect on pedohebephilic and pedophilic interests (Table 4-2). Restricting the analysis to treatment effect for the samples' highest response to stimuli depicting children produced similar findings (g = 0.34 [0.19, 0.48], Q = 1.7, I^2 = 0, k = 3, N = 187). Comprehensive programs had a small, positive effect on increasing teleiophilic interest. These treatment effects were not clinical or statistically greater than the natural history benchmark (Appendix D, Tables D-1 and D-2).

4.3.2.4 Eye movement and desensitization reprocessing. Two studies reported pre- to posttreatment changes over eye movement and desensitization reprocessing interventions (EMDR). These studies found significant positive change in pedohebephilic interests (g = 0.64, [0.14, 1.14], Q = 0.1, n = 13). This treatment effect was clinically larger than the natural history benchmark, but this effect was not statistically significantly greater than the natural history benchmark (p > .05).

4.3.3 Effect of Specific Behavioural Interventions

4.3.3.1 Pedohebephilic interests. Olfactory aversion showed a large, significant effect on pedohebephilic interest (g = 1.35 [0.57, 2.14], Q = 0.1, I^2 = 0, k = 2, N = 15). Moderate and significant effects were also found for covert and vicarious sensitization (g = 0.65 [0.407,0.891], Q = 0.3, I^2 = 0, k = 2, N = 75) and satiation (g = 0.76 [0.54, 0.99], Q = 9.6, I^2 = 58.3, k = 5, N = 89). Primary studies also reported on the effectiveness of combinations of behavioural interventions, which were grouped according the conditioning principles informing the intervention (e.g., positive reinforcement or extinction-based interventions). Moderate and

significant effects were found for combined positive reinforcement and extinction-based interventions (g = 0.60 [0.21, 1.00], Q = 5.2, I^2 = 55.4, k = 6, N = 93) and combined aversion and extinction-based interventions (g = 0.63 [0.45, 0.88], Q = 4.2, I^2 = 80.8, k = 2, N = 40). These two treatment effects were clinically and statistically greater than the natural history benchmark for pedohebephilic interest (g_{cv} = 0.60 and 0.57, $p < .05$; Appendix D, Table D-3). Small effects were found for combined signalled punishment and biofeedback (g = 0.39 [0.08, 0.70], Q = 0.9, I^2 = 0, k = 2, N = 150) and when positive reinforcement, aversion, and extinction interventions were combined (g = 0.19 [0.09, 0.29], Q = 9.4, I^2 = 89.3, k = 2, N = 410).[15] These two effects were not significantly greater than the natural history benchmark ($p > .05$)

4.3.3.2 Pedophilic interest. Two primary studies report ESs for satiation and found a large, significant effect (g = 1.08 [0.72, 1.45], $Q < 0.1$, I^2 = 0.9, k = 2, N = 33). Three single case reports also examined satiation and when these cases were included, the aggregate ES was g = 1.12 [0.76, 1.47]). The ES for satiation interventions was clinically and statistically greater than the natural history benchmark for pedohebephilic interest (g_{cv} = 0.78, $p < .01$; Table D-3). Small ES were found across two studies that combined aversion and extinction-based interventions (g = 0.30 [0.04, 0.55], $Q < .01$, I^2 = 0, k = 2, N = 58), which were not greater than the natural history benchmark ($p > .05$).

4.3.3.3 Teleiophilic interest. Studies reporting the effect of individual behavioural interventions on teleiophilic interests found non-significant ES for olfactory aversion (g = 0.12 [–0.35, 0.59], Q = 0.2, I^2 = 0, k = 2, N = 14), directed masturbation (g = –0.30 [–0.66, 0.05], Q = 4.7, I^2 = 78.7, k = 2, N = 33), and satiation (g = –0.03 [–0.28, 0.23], Q = 0.1, I^2 = 0, k = 2, N =

[15] Studies that combined all three forms of behavioural intervention reported on the effects of comprehensive treatment programs, while the other intervention types reported in this subsection come from studies that focused specifically on behavioural interventions.

17). Studies reporting combinations of behavioural interventions found small ES for aversion and extinction-based interventions (g = 0.23 [–0.05, 0.50], Q = 0.1, I^2 = 0, k = 2, N = 58) and positive reinforcement, aversion, and extinction-based interventions combined (g = 0.20 [0.10, 0.29], Q = 0.2, I^2 = 0, k = 2, N = 432).

4.3.4 Effect of Interventions in Sexual Offenders Against Children Subgroups

The effect of behavioural interventions in different subgroups of sexual offenders against children is examined according to the relationship of the offender to the victim(s), the gender of the victim(s), and the age of the offender. Behavioural interventions showed significant ES for pedohebephilic interest across all sexual offender against children subgroups examined (Table 4-3). For incest offenders, the treatment effect was clinically, but not statistically greater than the natural history benchmark (Table C-4). The treatment effects in the other offender subgroups were clinically and statistically greater than then natural history benchmark (p < .05). In contrast, behavioural interventions showed little effect for increasing teleiophilic interest (Table 4-3).

4.3.5 Treatment Effects across Taxon Membership

4.3.5.1 Pedohebephilic interests. Taxon membership was a significant moderator when all treatments were combined in the analysis and when behavioural and comprehensive treatments were combined (Table 4-4). Only aggregate effects in Taxon 3 were associated with decreases in pedohebephilic interest that were clinically and statistically greater than the natural history benchmarks (all ps < .01; Table C-5). There was not a significant difference in treatment effect across taxa when only behavioural intervention studies were considered.

4.3.5.2 Pedophilic interests. Taxon membership was a significant moderator when all treatments were combined in the analysis, when behavioural and comprehensive treatments were combined, and when analyses were restricted to behavioural interventions (Table 4-4). Only

treatments provided to samples classified in Taxon 3 was associated with decreases in pedophilic interest that were clinically and statistically greater than the natural history benchmarks (all *p*s < .01; Table C-5).

4.3.5.3 Teleiophilic interests. Taxon membership was a significant moderator when all treatments were combined in the analysis and when behavioural and comprehensive treatments were combined (Table 4-4). None of the treatment effects for teleiophilic interests were different from the natural history benchmarks.

4.3.6 Normative Comparisons

The two one-sided test procedure was conducted for pedohebephilic and pedophilic interest when all treatment types were combined, for combined behavioural treatments, and for pharmacological treatments. Teleiophilic interests were not examined, given that treatments had little effect on increasing interest in adults. The posttreatment PFE and index scores were equivalent to normative group data for pedohebephilic and pedophilic interests when all treatments were combined (Table 4-5). When *z*-score-based data were used, the posttreatment scores were not equivalent to normative data, indicating that posttreatment scores remained elevated. A similar pattern was observed for samples receiving behavioural interventions, with the exception that PFE data did not show equivalence and posttreatment scores remained elevated relative to normative data. Men receiving pharmacological treatments for pedohebephilic interests were not equivalent to normative men at posttreatment.

4.4 Discussion

Our meta-analysis of interventions for pedohebephilic interests demonstrated a “dodo bird” effect of sorts, consistent with treatments for other mental disorders (Luborsky et al., 2002). Diverse intervention methodologies demonstrated moderate to large reductions in

pedohebephilic arousal. The common thread for most of these interventions is helping the individual develop a skill set to: 1) attenuate pedohebephilic arousal, through pharmacological means or through conditioning by pairing arousal to children with a noxious odor, highly aversive imagined consequence, or boredom; 2) develop strategies to control that arousal such as via cognitive/behavioural techniques and/or reduction of serum testosterone; and/or 3) increase the interest in, or normalize, arousal to teleiophilic stimuli. The meta-analysis supported the former two propositions but support for the third was generally lacking in the available literature. The results transcended different age attractions and the age of offenders or subgroup the offender belonged to (i.e., incest or extrafamilial child victims). Few men are exclusively pedohebephilic; extant literature demonstrates that about a third to half of sexual offenders against children have pedohebephilic interests which usually coincide with some form of teleiophilic (i.e., age appropriate) interest, whether that be preferential or not (Seto, Lalumière, & Kuban, 1999). The motivation and ability to increase something that may already be there might be more difficult to detect.

The analyses grouping samples into three taxa suggest that, taking into account the sample sizes in each taxon across studies, between 67% and 86% of sexual offenders who underwent interventions tailored to reduced pedohebephilic interest did not show gains above natural processes. If these results reflect current clinical practice, this suggests a majority of individuals are undergoing treatment from which they do not benefit and represent a waste of therapeutic resources. By contrast, those men with the highest levels of pedohebephilic interest, who have the most room to change (i.e., taxon 3), demonstrated the most substantive changes in arousal and hence reductions in risk for future sexual contact with a child. The findings presented here reinforce the necessity of the Need principle in the risk-need-responsivity model of offender

rehabilitation: tailored interventions should only be provided to men who exhibit a problem with a risk factor (Bonta & Andrews, 2017).

This dodo bird effect does not extend to comprehensive treatment programs. One explanation is that these programs treated men with low levels of pedohebephilic interest. Indeed, most samples that underwent a comprehensive treatment program were classified into Taxon 1 in moderator analyses. Additionally, comprehensive programs were more likely to include a combination of behavioural treatment types, which may be less effective than using single behavioural interventions and which may also indicate less experimental control regarding what kinds and doses of treatment individual men were receiving. Comprehensive programs typically last a lengthy period of time, given the range of psychosocial issues addressed in these programs. Such a long period between test and retest may have also reduced detected treatment effects. At present, it is generally unclear if the ineffectiveness of comprehensive programs is due to providing services to men without pedohebephilic interests, or potentially due to the treatment approach itself or inappropriate combining of specialized interventions. However, there are compelling reasons for why improving intimacy and providing sexual education are important aspects of treatment, if reducing sexual recidivism is the aim of treatment (Marshall et al., 2009).

The presents findings have implications for the recent and ongoing debate regarding the flexibility of men's sexual interests in children (see Cantor, 2018; Fedoroff, 2018). However, an important limitation to the evidence reviewed here is that showing changes on phallometric testing is likely best understood as a change in a person's ability to monitor and manage their arousal as opposed to representing a shift in sexual orientation. Establishing that interventions are capable of shifting sexual orientation, which would include changes in sexual behaviour,

emotional and romantic attractions, sexual fantasy, sexual identity, and sexual arousal, is beyond the scope of the reviewed evidence base. While this limitation attenuates the conceptual implications of the findings, it is important to underscore that men who experience pedohebephilic interests and have committed sexual offences have, by definition, demonstrated an inability to control and manage their sexual arousal towards children, at least during the commission of their offence(s). Improving the ability to control arousal in everyday life is likely an important aspect of managing risk when released into the community; indeed, research suggests that men find behavioural interventions to be helpful in this regard (Milner, 2016).

4.4.1 Limitations and Future Directions

There are several noteworthy limitations to discuss. The most notable is that the use of no-treatment comparison groups and randomization was absent from most studies included. This results in most studies having an unknown or critical risk of bias. We made efforts to ameliorate the likelihood that natural processes might completely account for the findings, specifically using natural history benchmarks and posttreatment normative comparisons. However, we could not apply rigorous inclusion criteria to established higher quality natural history benchmarks. This limits the confidence we have in the benchmarking results and that treatment effects were due to intervention, and not natural processes.

A further limitation was that the *k* for some effect sizes was small and ranged from 2 to 11 studies. This concern is somewhat offset by limited effect size heterogeneity across most analyses (i.e., measures heterogeneity tended to be small in magnitude or not significant), and a high level of consistency between fixed effects and random effects analyses, both of which support the stability of findings and confidence in conclusions.

To address both concerns, research that uses randomized no-treatment or waitlist control group designs is needed to establish the efficacy of interventions. Such designs, using samples of men who display moderate to high levels of pedohebephilic interest, are currently feasible and are desperately needed. Given the centrality of pedohebephilic interest to preventing sexual offending, this is a priority for future research.

One limitation of the current state of psychological interventions for pedohebephilic interest is that technical innovation has stagnated for decades, with most studies included being published prior to 1995. Innovations in the treatment of pedohebephilic interests are needed, in order to keep abreast the growing understanding of the psychophysiology of human sexuality (Janssen, 2007), and possible, given the developing understanding of the influences of learning on sexuality (Hoffman, 2017). Another reason for technical innovation is that aversive interventions may not be acceptable to some clients and the negative side-effects of these interventions are unknown.

There is a need for the field to examine whether positive treatment change on indices of sexual interest predict reductions in sexual recidivism. Only two phallometric studies in the present sample of studies examined such associations; however, a much larger literature exists examining within treatment change in dynamic risk factors and its association with sexual offending, and a synthesis of this work will be forthcoming. Developing an empirical understanding of the relationship between change in pedohebephilic interests and sexual recidivism will improve our ability to target interventions and will establish pedohebephilic interests as a psychologically meaningful risk factor (Mann et al., 2010).

4.4.2 Conclusion

At present, there are no meta-analytic reviews of the effectiveness of interventions for pedohebephilic interests. The present meta-analysis represents a unique contribution that collects the current understanding of treatment effects and extends that knowledge in novel directions. The present results provide much reason for optimism in terms of helping men convicted of sexual offending manage their sexual interest in children. Most behavioural and pharmacological interventions were found to be associated with reductions in pedohebephilic arousal, especially for those men who showed high levels of such arousal. However, there was little evidence that these interventions can increase sexual interest in adults and there was little evidence that comprehensive programs are effective.

The clinical implications are considerable. These findings provide clinicians with an evidence guide and recommendations on which clients to offer specialized services to, what interventions may work with their clients, and the amount of change they and their clients can expect to make over the course of treatment. In short, men and youth are capable of managing pedohebephilic arousal when trained in behavioural techniques to manage that arousal. And when examined, the arousal attenuation results in levels of arousal on par with that displayed by non-pedohebephilic individuals. We do not know the long-term effects of these interventions or their staying power. But in principle, the capacity to control or inhibit arousal is a portable skill that can be pivotal to risk management and the prevention of sexual offending against children. Future research using more rigorous methodologies and technical innovations will further advance the field by establishing the efficacy of treatments, expanding the intervention choices for clinicians to offer their clients, and address the important relationship of within treatment change and sexual recidivism.

Table 4-1

Characteristics of studies included in the meta-analysis

Source	Treatment *n*	Subgroups	Treatment modality	Specific interventions	Risk of Bias
Aldridge	42	SOC, SOC-E	Comprehensive	Satiation, olfactory aversion, covert sensitization, thought-stopping, impulse counting	Critical
Bedard	46	SOC-MV, SOC-FV	Comprehensive	Satiation, olfactory aversion, cognitive restructuring	Critical
Bradford & Pawlak	17	SOC	Pharmacological	Cyproterone acetate	Critical
Cooper et al.	7	SOC	Pharmacological	Medroxyprogesterone acetate, Cyproterone acetate	Critical
Crolley et al.	16	SOC	Behavioural	Masturbatory reconditioning, covert sensitization	Critical
Dohrmann	3	SOC	EMDR	--	Critical
Gray	25	SOC	Behavioural	Minimal arousal conditioning, satiation	Serious
Hunter & Santos	27	JSOC-MV, JSOC-FV	Behavioural	Satiation, covert sensitization	Critical
Johnston et al.	10	SOC-E	Behavioural	Masturbatory reconditioning	Critical
Jones	311	SOC	Comprehensive	Satiation, covert sensitization, directed masturbation	Critical
Kaplan et al.	15	JSOC	Behavioural	Satiation	Critical
Lang	3	SOC	Behavioural	Satiation	Critical
Marques et al.	171	SOC-E	Comprehensive	Masturbatory reconditioning, satiation, olfactory aversion	Critical
Marshall & Barbaree	68	SOC-E, SOC-I	Behavioural	Masturbatory reconditioning, olfactory aversion, electrical aversion	Critical
Marshall	12	SOC-EFV	Other	Self-esteem focused therapy	Critical
Johnston	30	SOC-E	Behavioural	Satiation, Directed masturbation	Critical
Quinsey et al.	18	SOC	Behavioural	Biofeedback and signalled punishment	Critical
Ricci et al	10	SOC	EMDR	--	Serious
Rice et al.	50	SOC-E	Behavioural	Biofeedback, signalled punishment shock	Serious
Schober et al.	5	SOC	Pharmacological,	LHRH, CBT	Critical

			CBT		
Weinrott et al.	35	JSOC	Behavioural	Vicarious sensitization	Moderate
Hunter & Goodwin	27	JSOC	Behavioural	Satiation	Critical
Clift et al.	106	JSOC	Behavioural, CBT	Covert sensitization + Thought-stopping, positive self-talk, impulse control	Critical
Single case designs					
Alford et al.	1	SOC-FV	Behavioural	Masturbatory extinction	Critical
Earls & Castonguay	1	SOC-E	Behavioural	Olfactory reconditioning	Critical
Foote & Laws	1		Behavioural	Masturbatory reconditioning	Critical
Harbert et al.	1	SOC-I	Behavioural	Covert sensitization	Critical
Kremsdorf et al.	1	SOC-FV	Behavioural	Directed masturbation	Critical
Laws	1	SOC-MV	Behavioural	Biofeedback	Critical
Laws	1	SOC-MV	Behavioural	Olfactory reconditioning	Critical
Lee	4	SOC-I, SOC-E	Behavioural	Olfactory reconditioning	Critical
Levin et al.	1		Behavioural	Covert sensitization, valeric acid	Critical
Marshall & Barbaree	2	SOC-FV	Behavioural	Satiation, aversion	Critical
Marshall	2	SOC-EFV	Behavioural	Masturbatory reconditioning, satiation; Satiation, aversion	Critical
Marshall	1	SOC-EFV	Behavioural, CBT	Olfactory aversion, directed masturbation, cognitive restructuring, cognitive distortions	Critical
Marshall	1	SOC-EMV	Behavioural	Covert association	Critical
Plaud & Gauthier	1	SOC-EMV	Behavioural	Covert sensitization	Critical
Rea et al.	1		Behavioural	Covert sensitization	Critical
Stava et al	1		Behavioural	Covert sensitization	Critical
VanDeventer & Laws	2	SOC-MV	Behavioural	Masturbatory reconditioning	Critical
Wincze et al.	3	SOC-MV, SOC-FV	Pharmacological	Medroxyprogesterone acetate	Critical

Note. CBT = Cognitive behavioural therapy; EMDR = Eye movement desensitization and reprocessing; JSOC = Juvenile sexual offenders against children; LHRH = Luteinizing hormone-releasing hormone agonist; SOC = sexual offenders against children; SOC-E = sexual offenders with unrelated child victims; SOC-FV = sexual offenders with female child victims; SOC-I = sexual offenders against related children; SOC-MV = sexual offenders with male child victims.

Table 4-2

Effects of interventions on pedohebephilic, pedophilic, and teleiophilic interests

	Fixed-effect		Random-effects		*Q*	I^2	*N* (*k*)
	g	[95% CI]	*g*	[95% CI]			
Behavioural Treatments							
Pedohebephilic interest	0.612	[0.488, 0.737]	0.633	[0.469, 0.798]	16.90	28.98	408 (13)
with single case designs	0.657	[0.535, 0.780]	0.767	[0.541, 0.994]	33.45**	61.14	432 (14)
Pedophilic interest	0.778	[0.515, 1.040]	0.747	[0.306, 1.188]	7.90	49.37	56 (5)
with single case designs	0.840	[0.583, 1.098]	0.927	[0.408, 1.445]	13.47	62.88	64 (6)
Teleiophilic interest	–0.103	[–0.286, 0.079]	–0.088	[–0.352, 0.175]	8.99	44.36	75 (6)
Pharmacological Treatments							
Pedohebephilic interest	0.648	[0.305, 0.991]	0.648	[0.305, 0.991]	1.06	0.00	32 (4)
Comprehensive Treatments							
Pedohebephilic interest	0.202	[0.119, 0.284]	0.256	[0.109, 0.404]	11.36*	55.98	587 (6)
Pedophilic interest	0.122	[0.027, 0.216]	0.131	[0.022, 0.240]	2.28	12.13	428 (3)
Teleiophilic interest	0.196	[0.102, 0.290]	0.196	[0.102, 0.290]	0.22	0.00	474 (3)

Table 4-3

Meta-analyses of changes in pedohebephilic interest, pedophilic interest, and teleiophilic interest during behavioural treatments in sexual offenders against children subgroups

Subgroup	Fixed-effect		Random-effects		*Q*	I^2	*N* (*k*)
	g	[95% CI]	*g*	[95% CI]			
Pedohebephilic Interest							
SOC-E	0.759	[0.518, 1.000]	0.800	[0.470, 1.131]	3.12	35.97	83 (3)
SOC-I	0.551	[0.177, 0.924]	0.674	[–0.007, 1.356]	1.70	41.02	29 (2)
SOC-FV	0.578	[0.308, 0.847]	0.678	[0.123, 1.233]	1.47	32.14	57 (2)
Adult SOC	0.705	[0.551, 0.858]	0.854	[0.551, 1.157]	28.47**	64.88	321 (11)
Juvenile SOC	0.575	[0.373, 0.777]	0.591	[0.262, 0.920]	3.98	49.71	111 (3)
Teleiophilic Interest							
SOC-E	–0.484	[–0.814, –0.154]	–0.484	[–0.814, –0.154]	0.01	0.00	40 (2)
Adult SOC	–0.103	[–0.286, 0.079]	–0.088	[–0.352, 0.175]	8.99	44.36	75 (6)

Note. SOC = sexual offenders against children; SOC-E = sexual offenders with unrelated child victims; SOC-FV = sexual offenders with female child victims; SOC-I = sexual offenders against related children.

Table 4-4

Moderator analyses of treatment change as a function of taxon membership

Treatment Type	Fixed-effect		Random-effects		
	g	[95% CI]	*g*	[95% CI]	*Q-between*
Pedohebephilic Interests					
All treatments combined					44.45***
Taxon 1	0.121	[0.012, 0.229]	0.121	[0.012, 0.229]	
Taxon 2	0.436	[0.263, 0.608]	0.434	[0.247, 0.621]	
Taxon 3	0.741	[0.592, 0.890]	0.889	[0.626, 1.151]	
Behavioural + Comprehensive					41.36***
Taxon 1	0.121	[0.012, 0.229]	0.121	[0.012, 0.229]	
Taxon 2	0.416	[0.237, 0.595]	0.404	[0.175, 0.633]	
Taxon 3	0.763	[0.597, 0.928]	1.004	[0.644, 1.364]	
Behavioural treatments					2.17
Taxon 1	0.305	[–0.170, 0.780]	0.305	[–0.170, 0.780]	
Taxon 3[a]	0.684	[0.514, 0.855]	0.773	[0.514, 1.033]	
Pedophilic Interest					
All treatments combined					32.95***
Taxon 1	0.091	[–0.018, 0.199]	0.091	[–0.018, 0.199]	
Taxon 2	0.297	[–0.001, 0.595]	0.297	[–0.001, 0.595]	
Taxon 3	0.952	[0.667, 1.227]	1.065	[0.535, 1.596]	
Behavioural + Comprehensive					37.26***
Taxon 1	0.091	[–0.018, 0.199]	0.091	[–0.018, 0.199]	
Taxon 2	0.297	[–0.001, 0.595]	0.297	[–0.001, 0.595]	
Taxon 3	1.181	[0.846, 1.516]	1.257	[0.784, 1.730]	
Behavioural treatments					9.00**
Taxon 1	0.293	[–0.180, 0.767]	0.293	[–0.180, 0.767]	
Taxon 3	1.181	[0.846, 1.516]	1.257	[0.784, 1.730]	
Teleiophilic Interest					
All treatments combined					14.08**
Taxon 1	0.211	[0.102, 0.321]	0.211	[0.102, 0.321]	
Taxon 2	0.168	[–0.004, 0.340]	0.168	[–0.004, 0.340]	
Taxon 3	–0.196	[–0.381, –0.010]	–0.207	[–0.427, 0.014]	
Behavioural + Comprehensive					13.95**
Taxon 1	0.211	[0.102, 0.321]	0.211	[0.102, 0.321]	
Taxon 2	0.168	[–0.004, 0.340]	0.168	[–0.004, 0.340]	
Taxon 3	–0.192	[–0.395, 0.011]	–0.211	[–0.499, 0.077]	
Behavioural treatments					2.83
Taxon 1	0.278	[–0.230, 0.786]	0.278	[–0.230, 0.786]	
Taxon 3	–0.192	[–0.395, 0.011]	–0.211	[–0.499, 0.077]	

[a] The single case design studies' aggregate effect size was identified as an outlier in this analysis. The aggregate ES reported here does not include the effect size from the single case design studies.

**$p < .01$. ***$p < .001$.

Table 4-5

Comparisons in pedohebephilic arousal pre and posttreatment between SOC and a normative comparison sample

	Normative sample			SOC sample			Equivalent
Data Type	*M*	*SD*	*k* (*n*)	*M*	*SD*	*k* (*n*)	
All treatments combined							
Pedohebephilic							
PFE	24.56	34.10	6 (162)	27.83	25.06	12 (255)	Yes
Index	0.65	0.66	4 (98)	0.69	0.50	11 (223)	Yes
z-score	0.03	0.81	2 (75)	0.58	0.62	1 (3)	No
z-score index	–1.18	1.05	3 (145)	–0.20	1.16	3 (102)	No
Pedophilic							
PFE	14.99	17.95	2 (36)	15.63	17.14	5 (87)	Yes
Index	0.68	0.99	(36)	0.34	0.31	2 (19)	Yes
z-score	0.03	0.81	2 (75)	0.58	0.61	1 (3)	No
z-score index	–0.98	1.04	1 (112)	–0.23	1.20	2 (78)	No
Behavioural treatments							
Pedohebephilic							
PFE	24.56	34.10	6 (162)	34.54	27.47	10 (186)	Yes
Index	0.65	0.66	4 (98)	0.70	0.53	10 (211)	Yes
z-score	0.03	0.81	2 (75)	0.58	0.62	1 (3)	No
z-score index	–1.18	1.05	3 (145)	–0.07	0.37	1 (3)	No
Pedophilic							
PFE	14.99	17.95	2 (36)	26.68	21.48	4 (45)	No
Index	0.68	0.99	2 (36)	0.34	0.31	2 (19)	Yes
z-score	0.03	0.81	2 (75)	0.58	0.61	1 (3)	No
z-score index	–0.98	1.04	1 (112)	–0.07	0.23	1 (3)	No
Pharmacological treatments							
Pedohebephilic							
PFE	14.99	17.95	2 (36)	20.45	14.49	3 (17)	No

Note. PFE = percent full erection; *k* = number of studies included; *SD* = standard deviation.

CHAPTER 5

CONVERGENT AND PREDICTIVE ASSOCIATIONS OF THREE MEASURES OF PEDOPHILIC INTEREST

5.1 Introduction

Pedophilic interest connotes a sexual attraction to prepubescent children and is a main risk factor for the initiation of sexual offending against children (Seto, 2018). Meta-analytic research has identified pedophilic interest as having one of the strongest predictive relationships with sexual recidivism in men convicted of sexual offenses against children (Hanson & Morton-Bourgon, 2005; McPhail et al., 2017). Given the relative theoretical and clinical importance of pedophilic interest, most treatment programs provide specialized interventions to help men manage their arousal to children (McGrath et al., 2010) and a majority of risk instruments that include dynamic risk factors assess for pedophilic interest (or broader paraphilic interests) to evaluate risk for future sexual violence (Hanson & Morton-Bourgon, 2009).

5.1.1 Measures of Pedophilic Interest

Validating measures of pedophilic interests is an important task for applied forensic research. Because pedophilic interest is theorized to contribute to the initiation of sexual offending against children, valid measures are expected to differentiate men who have and men who have not committed sexual offences against children. Further, one of the main tasks of clinical forensic work is the evaluation of sexual recidivism risk, valid measures of pedophilic interest should have a predictive association with future sexual offending by men who have been convicted of sexual offences. Several measures of pedophilic interest have been developed and tested in samples of men with sexual offence histories. Here we provide an overview of three measures of pedophilic interest used with men with sexual offence histories.

5.1.1.1 Phallometric tests of pedophilic interest. Phallometric testing for sexual arousal to children is frequently employed as a measure of pedophilic interest in men. There is variety in the procedures used during phallometric tests (i.e., differences in stimulus presentation modality; method of measuring changes in penile arousal; methods of transforming raw phallometric data; etc.) and this lack of standardization has presented a problem for the validation of phallometric tests for pedophilic interest (Marshall & Fernandez, 2000). However, recent meta-analytic research has shown that most phallometric procedures differentiate sexual offenders against children from other samples and phallometric test scores from most procedures predict sexual recidivism (McPhail et al., 2017). Other meta-analytic research indicates that phallometric test scores are related to viewing time measures of pedophilic interests (viewing time r = .25; Schmidt, Babchishin, & Lehmann, 2017). However, few examinations have established whether pedophilic interest, as measured by phallometric testing, has incremental predictive power beyond measures of static risk. Such validity of research would provide evidence that phallometric test scores provide unique prediction of sexual recidivism to established actuarial risk assessment tools.

5.1.1.2 Screening Scale for Pedophilic Interest (SSPI). The SSPI is a brief screening instrument to assess for pedophilic interest (Seto & Lalumière, 2001). Validity research has generally found that the SSPI has a significant, if small to moderate, relationship with sexual recidivism in samples of men convicted of sexual offences (Eher, Olver, Heurix, Schilling, & Rettenberger, 2015; Helmus, Ó Ciardha, & Seto, 2014; Seto, Harris, Rice, & Barbaree, 2004; Seto, Sandler, & Freeman, 2017). However, other research has not found the SSPI to predict sexual recidivism and to not add incremental predictive power beyond static risk, as measured by Static-99R (Canales, Olver, & Wong, 2009; Moulden, Firestone, Kingston, & Bradford, 2009;

Seto et al., 2017), or pedophilic interest, as measured by the DSM (Eher et al., 2015). Other validity research indicates that the SSPI is associated with phallometric (r = .34; Seto & Lalumière, 2001), viewing time (r = .21; Schmidt et al., 2017), and implicit association test measures of pedohebephilic interests[16] (r = .28; Babchishin, Hermann, & Nunes, 2013).

5.1.1.3 Violence Risk Scale-Sexual Offense version (VRS-SO). The VRS-SO is a sexual offense risk assessment instrument that measures static risk factors, dynamic risk factors, and treatment change of the dynamic risk factors (Wong, Olver, Nicholaichuk, & Gordon, 2003, 2017). Factor analytic research has found three correlated dimensions underlaying the VRS-SO dynamic risk items: Sexual Deviance, Criminality, and Treatment Responsivity factors (Beggs & Grace, 2010; Olver, Wong, Nicholaichuk, & Gordon, 2007; Olver & Eher, 2019; Olver, Neumann, Kingston, Nicholaichuk, & Wong, 2018). The Sexual Deviance factor comprises five items: Sexually Deviant Lifestyle, Sexual Compulsivity, Offense Planning, Sexual Offending Cycle, and Deviant Sexual Preference; this lattermost item constitutes a measure of deviant sexual interest that encompasses sexual attraction to children. Validity research has found that men with unrelated child victims and men with unrelated child and adult victims score higher on the Sexual Deviance factor compared with men with adult victims and men with related child victims (Canales et al., 2009; Olver & Wong, 2006), and that this domain predicts sexual recidivism (Beggs & Grace, 2010; Olver et al., 2007). A significant, positive relationship has also been found between Sexual Deviance factor scores and other measures of pedophilic interest, such as phallometric tests of pedophilic interest and the SSPI (Canales et al., 2009).

5.1.2 Taxonicity and Dimensionality of Pedophilic Interests

[16] *Pedohebephilic interests* denotes a sexual attraction to prepubescent and/or early pubescent children.

Recent taxometric studies have produced support for three separate methods for modelling the distribution of pedophilic interest in men (McPhail, Olver, Brouillette-Alarie, & Looman, 2018; Schmidt, Mokros, & Banse, 2013; Stephens, Leroux, Skilling, Cantor, & Seto, 2017), which in turn, have implications for the measurement and conceptualization of pedophilic interest. Taxometric analysis is a set of procedures that empirically tests whether a psychological construct is better characterized as a latent dimension or a latent taxon (or comprised of multiple latent taxa; Ruscio, Haslam, & Ruscio, 2006; Waller & Meehl, 1998). Pedophilic interest, across these recent taxometric analyses, has been found to be better characterized as a dimension (Stephens et al., 2017), a taxon (i.e., dichotomously distributed; Schmidt et al., 2013), and two ordered taxa (i.e., trichotomously distributed; McPhail et al., 2018).

There are notable limitations to each of the taxometric studies and further taxometric replication studies are required to further illuminate the latent structure of pedophilic interest. An additional way to further our understanding of the latent structure is to test the validity of different latent structural models to evaluate which model(s) are supported. The underlying assumption of such validity research, and indeed of taxometric research itself (Ruscio et al., 2006), is that the best fitting latent structure will better model observable scores on the construct of interest and produce more valid results. Modelling observable data to more closely fit the latent structure of pedophilic interest should lead to improved validity of scores on tests of pedophilic interest.

Past research has provided indirect or partial tests of the predictive validity of different latent structural models of pedophilic interest. These studies provide indirect tests because latent structure was either not explicitly considered by the researchers, or partial tests because the researchers only tested one or two latent structural models. Research operationalizing pedophilic

interests using a dichotomous model has typically found that pedophilic interest is not predictive of sexual recidivism in men convicted of sexual offences against children (Eher, Rettenberger, Matthes, & Schilling, 2010; Moulden et al., 2009; Stephens et al., 2017; Wilson, Abracen, Looman, Pichea, & Ferguson, 2011). These studies model pedophilic interests by grouping a sample into pedophilic and non-pedophilic men. Research using a trichotomous operationalization of pedophilic interests finds preferential and exclusive pedophilic interest to be a strong predictor of sexual recidivism (Biere, 1998; Eher et al., 2010; Eher et al., 2015; McPhail et al., 2018). In these studies, pedophilic interest is operationalized by either grouping samples into three groups (i.e., non-pedophilic, non-preferentially/non-exclusively pedophilic, and preferentially/exclusively pedophilic; refs) or into two groups (i.e., non-pedophilic or non-preferentially/non-exclusively pedophilic and preferentially/exclusively pedophilic). Most studies in a recent meta-analysis modelled pedophilic interest continuously and found overall support for a relationship between pedophilic interest and sexual recidivism (McPhail et al., 2017). More recent research using a continuous model did not find such support (Stephens et al., 2017). This body of research, even though modelling underlying latent structure in pedophilic interest and testing these models was not an explicit aim of the studies, provides support for a trichotomous model for pedophilic interest and some support for a dimensional model.

5.1.3 Present Study

Validating measures of pedophilic interests is an ongoing task for applied research. To contribute to the ongoing validation of measures of pedophilic interest, the present study examines the convergent and predictive validity of three measures of pedophilic interest by (a) replicating and extending the findings reported in Canales et al. (2009) using a larger sample than was available to those researchers and (b) subjecting measures of pedophilic interest to a

more severe validity test by assessing whether the measures of pedophilic interest predict recidivism over-and-above an established actuarial static risk instrument. Given the emerging, and conflicting, results regarding the latent structure of pedophilic interest, the present study will test the validity of latent structural models derived from recent taxometric analyses of pedophilic interests. This aim will provide a test for which latent structure provides the most optimal operationalization of pedophilic interest.

5.2 Method

5.2.1 Participants

The present sample included 261 men convicted of sexual offenses who underwent assessment and treatment services at the Regional Psychiatric Centre in Saskatoon, Canada. The men had participated in the Clearwater Program, which was a high-intensity treatment program for federally incarcerated men convicted of sexual offenses. The sample comprised 91 men who had been convicted of sexual offences against children under the age of 14 and 170 men who had been convicted of sexual offences against individuals over the age of 14. The sample was 35.5 years of age at release (*SD* = 9.8) and had 1.0 (*SD* = 1.5) prior convictions for sexual offenses and 1.3 prior convictions for non-sexual violent offenses (*SD* = 1.9).

5.2.2 Measures

5.2.2.1 Phallometric testing. Phallometric testing was conducted using a mercury-in-rubber strain gauge to measure changes in men's penile circumference during the presentation of 16 slides of naked or partially naked individuals. There were different age categories of individuals depicted in the slides (i.e., 5 to 10-year-old children; 12 to 15-year-old children; and 18 and older adults) across both sexes. In the present study, percent full erection responses to slide depicting 5 to 10-year-old children were used as a measure of pedophilic interest. The

average percent full erection to prepubescent children overall, male child, and female children were computed and used in the analyses. The current sample underwent phallometric testing as part of a routine assessment prior to beginning treatment. For further details of the phallometric testing procedure, see Canales et al. (2009).

5.2.2.2 Screening Scale for Pedophilic Interest (SSPI). The SSPI is a four-item measure of sexual interests in children based on victim characteristics in men convicted of sexual offences against children (Seto & Lalumière, 2001). Total scores on the SSPI range from 0 to 5. The SSPI is related to phallometrically assessed sexual interest in children (r = .27, n = 145; Seto et al., 2004) and to sexual recidivism (AUC = .69, n = 130, Seto et al., 2004; AUC = .62, n = 365, Helmus, et al., 2015).

5.2.2.3 Violence Risk Scale-Sexual Offense Version (VRS-SO). The VRS-SO is a clinician-rated risk assessment tool designed to predict sexual recidivism and monitor treatment change in adult males convicted of sexual offences (Wong et al., 2003, 2017). Both static and dynamic risk factors are measured on the VRS-SO, with the dynamic risk factors being represented by three underlying dimensions, including a Sexual Deviance factor, as previously noted. All items are rated on a 4-point ordinal scale, with scores ranging from 0 to 3 (i.e., from a risk factor being absent to being present for an individual). Scores on the Sexual Deviance factor range from 0 to 15 and the items can be rated across multiple time points (e.g., pre and posttreatment). The Sexual Deviance factor score has been found to have acceptable interrater reliability for the pretreatment and posttreatment scores (intraclass correlation coefficient [ICC] = .72 and .73, respectively; Olver et al., 2007). VRS-SO Sexual Deviance factor ratings were obtained from Olver et al. (2007).

5.2.2.4 Static-99R. Static-99R is a 10-item static empirical actuarial risk assessment tool designed to assess risk for sexual recidivism in adult males adjudicated for sexual offenses (Hanson & Thornton, 2000; Helmus, Thornton, Hanson, & Babchishin, 2012). Items on Static-99R are readily scored using information from archival sources (e.g., criminal records). The range of scores on Static-99R is –3 to 12, with higher scores indicating higher risk. Static-99R has demonstrated robust predictive accuracy for sexual recidivism (average AUC = .70 in a recent meta-analysis; *N* = 8,055; Helmus, Hanson, Thornton, Babchishin, & Harris, 2012). File reviews (*n* = 88) have found high reliability of Static-99 scores between community supervision officers and expert ratings (ICC = .91; see Hanson, Harris, Scott, & Helmus, 2007). Static-99R scores were obtained from converted Static-99 ratings from Olver et al. (2007).

5.2.2.5 Sexual recidivism. Sexual recidivism in the sample was defined as a conviction for a new offense incurred post release that was sexual in nature or was rated to be sexually motivated after a review of offense details (e.g., a homicide offence was judged to be a sexual homicide based on review of police documents). Sexual recidivism was retrieved from the Canadian Police Information Centre and was coded as either present or absent. The average postrelease follow-up time for the sample was 12.2 years. The overall sexual recidivism rate in the sample was 29.2%.

5.2.3 Analytic Plan

The following analyses were conducted to examine the convergent and predictive associations of the three aforementioned measures of pedophilic interest as follows.

5.2.3.1 Convergent validity in measures of pedophilic interest. Zero-order correlations between the measures of pedophilic interest were performed in the full sample and a sample restricted to men who had committed sexual offences against children under the age of 14.

Percentage of shared variance among the measures of pedophilic interest are also reported in text (i.e., percent shared variance = [r^2] * 100). Cohen's (1988) guidelines for the interpretation of correlation magnitudes between two continuously measured variables were employed in which values of .10, .30, and .50 correspond to small, medium, and large effects respectively.

5.2.3.2 Predictive validity in measures of pedophilic interest. To examine the association between the measures of pedophilic interest and sexual recidivism, the first set of analyses calculated the area under the receiver operating curve (AUCs) for the measures of pedophilic interests. An AUC represents the probability that for two randomly selected offenders, one who recidivated and one who did not recidivate, the measure of pedophilic interest will correctly classify the recidivist as having a higher score on the measure. AUCs of .56, .64, and .71 are considered to be small, moderate, and large effects (Rice & Harris, 2005).

Second, Cox regression survival analysis was employed to examine the prediction of sexual recidivism over time. Scores on each measure of pedophilic interest were initially entered individually into the model as a single predictor of sexual recidivism controlling for follow-up time. Cox regressions generate a hazard ratio (e^B), which represents the change in relative risk of the outcome (i.e., sexual recidivism) occurring for each one unit change on the predictor variable; values above 1.0 indicate a positive association. In subsequent steps, Static-99R was entered as a predictor followed by scores on a given measure of pedophilic interests to examine their unique and incremental associations with outcome. In addition to the individual e^B values assigned to model predictors evaluate significance, the change in the model (χ^2-change) indicates whether incremental improvement in the prediction of sexual recidivism has been obtained from the Static-99R only model.

5.2.3.3 Predictive validity in latent structural models of pedophilic interest. In order to test the performance of different latent structural models of pedophilic interest, phallometric test scores were used to measure the distribution of scores in ways that map onto the results of recent taxometric analyses. The phallometric tests scores were then entered as predictors in unique Cox regression analyses. In order to capture the models presented in recent taxometric analyses, the latent structural models used in the analyses were: a continuous, a dichotomous, and two trichotomous latent structural models of pedophilic interest. For the continuous latent structural model, percent full erection (PFE) phallometric test scores were entered a predictor into a Cox regression model. For the dichotomous model, the PFE cut score, provided by McPhail et al. (2018), that differentiated between the non-pedophilic and pedophilic taxa, was used to categorize the sample into two groups. This dichotomous predictor variable was then entered into a Cox regression model. For the first trichotomous model, two PFE cut scores, provided by McPhail et al. (2018), were used to categorize the sample into three separate groups. This three-group predictor variable was then entered into a Cox regression model. For the second trichotomous model, the percent full erection cut score that differentiated between the first and second pedophilic taxa in McPhail et al. (2018) was used to categorize the sample into two groups. This two-group predictor variable was then entered into a Cox regression model.[17] In composing each model, the average PFE score across female and male prepubescent child stimulus trials, the average PFE score across female child stimulus trials, and the average PFE score across male child stimulus trials were used in separate analyses. As with the above Cox

[17] This second method of operationalizing a trichotomous latent structural model was conducted because, on inspecting the sexual recidivism rates across the three taxa, McPhail et al. (2018) reported that the sexual recidivism rates for the teleiophilic taxon and non-preferentially pedophilic taxon were almost equal. Given this equality in the odds of the outcome occurring these two taxa, we would not expect a latent structural model that treats these two taxa as distinct to perform well in a Cox regression model. This was determined a prior and does not represent a post hoc analysis.

regression analyses, two series of analyses were conducted. In the first, each models' variable was entered as the sole predictor, and then subsequently with the Static-99R entered in the first step and followed by the latent structural model in the second step.

5.3 Results

5.3.1 Convergent validity of measures of pedophilic interest

Zero-order correlations between the measures of pedophilic interest are provided in Table 4-1.[18] The relationships between the phallometric responses were large and significant, for both the full sample and the sexual offenses against children (SOC) sample, and the amount of shared variance between phallometric measures of pedophilic interest ranged from 29% to 86%. The phallometric measures generally showed moderate and significant associations with the other measures of pedophilic interest and shared between 6% and 21% variance with these measures. The exception was the SSPI, which showed moderate and significant associations with phallometric measures in the full sample (6% to 21% shared variance), while in the SOC sample, the associations with the child and female child stimulus responses were small and non-significant (shared variance of 6% and <1%, respectively). In the SOC sample, there was a moderate and significant association between the SSPI and the phallometric measure of male-oriented pedophilic interest (shared variance = 17%); this relationship may be stronger owing to the SSPI having an item assessing the presence of male victims in the sexual offense history. The VRS-SO Sexual Deviance factor scores showed moderate and significant associations with the phallometric measures and SSPI.

5.3.2 Predictive validity in measures of pedophilic interest

[18] Given the skew in the phallometric scores Spearmen's correlations were run in addition to Pearson's correlations. When a non-parametric Spearman's coefficient was different from the Pearson's coefficient at the level |.10|, the non-parametric coefficient is reported in Table 4-1.

The phallometric measures of arousal to children and female children showed small to moderate and significant associations with sexual recidivism in both the full sample and the SOC sample (Tables 4-2 and 4-3). The phallometric measure of arousal to male children showed a small, non-significant association with sexual recidivism in the full sample, while the association was moderate and significant in the SOC sample. The SSPI had small, non-significant associations with sexual recidivism in both the full and more restricted samples. The VRS-SO Sexual Deviance factor scores showed small to large, significant associations with sexual recidivism. The pattern of results suggests that in the full sample, associations were small (i.e., all AUCs < .64), while the associations were moderate to large in the SOC sample (i.e., AUCs between .64 and .72). The same pattern was found when each measure of pedophilic interest was entered into separate Cox regression models.

In the incremental validity analyses controlling for Statc-99R score, the SSPI (both samples) and the phallometric measure of interest in boys (total sample), were not included given that they did not have significant bivariate associations with sexual recidivism. In the full sample, only VRS-SO Sexual Deviance posttreatment scores added incremental predictive power beyond Static-99R (Table 4-4). No other measures of pedophilic interest significantly and incrementally predicted sexual recidivism, although this interpretation may warrant consideration of possible Type II error, given the magnitude of the hazard ratios and several non-significant results at $p < .10$.

5.3.3 Predictive validity in latent structural models of pedophilic interest

Cox regression results for the association of phallometric test scores, using different latent structural models, with sexual recidivism in the total sample and sexual offender against children sample are presented in Table 4-5. Across interest in both sexes and interest in both female and

male children, the continuous model was significantly predictive of sexual recidivism, while the dichotomous model was not associated with sexual recidivism. The three-level trichotomous model predicted sexual recidivism, but only in the SOC sample and for interest in both sexes and male children. The two-level trichotomous model predicted sexual recidivism consistently in the SOC sample.

These significant predictive relationships were subjected to the more severe validity test to assess whether latent structural models of pedophilic interest add incremental predictive power when controlling for static risk (Table 4-6). The trichotomous model for sexual interest in both sexes of children combined, using two levels (i.e., preferentially pedophilic men vs. non-preferentially pedophilic and non-pedophilic men), was a significant predictor of sexual recidivism, after controlling for Static-99R, in the SOC sample. The preferentially pedophilic men had over three times the relative risk of sexually re-offending compared to the other SOC men in the sample, with the recidivism rates been markedly different (75% vs. 24%). The continuous model did not continue to predict sexual recidivism when controlling for Static-99R scores. When arousal to female children was examined, the same pattern of results emerged; whereas no latent structural model of arousal to male children predicted sexual recidivism when controlling for Static-99R scores.

5.4 Discussion

The present study examined validity in measures of pedophilic interest in a sample of incarcerated men. There was support for phallometric and VRS-SO measures as assessing similar constructs; there was less support for the SSPI as a measure of pedophilic interest, as this measure showed a somewhat more inconsistent pattern of relationships with the phallometric measures. A similar pattern of results emerged when the associations between the measures of

pedophilic interest and sexual recidivism were examined. Phallometric testing and the VRS-SO showed consistent predictive associations, while the SSPI was not significantly related to sexual recidivism.

The present results provide validity support to the VRS-SO Sexual Deviance factor, extending the results of Canales et al. (2009) in a larger sample. The moderate correlations between the VRS-SO Sexual Deviance scores and phallometric testing support this factor of the VRS-SO capturing pedophilic interests. A somewhat different pattern was observed for the relationship between the SSPI and phallometric testing. Specifically, the SSPI was moderately correlated to arousal to male children for the SOC sample. In contrast, the correlation between the SSPI and arousal to female children and overall arousal to children did not reach statistical significance. This nonsignificant result is not consistent with past research examining the relationship of the SSPI with phallometric testing (Seto & Lalumière, 2001); however, the overall magnitude of the relationship is relatively similar in both studies. In interpreting the absolute magnitude of these correlations with phallometric testing, the VRS-SO Sexual Deviance factor and the SSPI may be thought of as proxies for pedophilic interests. The SSPI provides behavioral proxies of pedophilic interests, while the VRS-SO Sexual Deviance factor itself assesses a broader range of risk-relevant constructs than pedophilic interest.

The VRS-SO Sexual Deviance scores consistently predicted sexual recidivism and when subjected to a severe validity test, the VRS-SO Sexual Deviance post-treatment score was associated with sexual recidivism after controlling for static risk. These results provide further evidence to the growing body of research supporting the validity of these factor scores as being meaningfully related to recidivism outcomes (Beggs & Grace, 2010; Canales et al., 2009; Olver & Wong, 2006). The SSPI was generally not associated with sexual recidivism. The existent

literature examining the predictive validity of the SSPI has produced mixed findings and the present finding contributes to the mixed support. However, given the state of the science regarding the predictive validity of the SSPI, a meta-analytic estimate may resolve this inconsistency.

Testing different methods of operationalizing latent structure of pedophilic interest, as measured by phallometry, showed that how pedophilic interests are measured and conceptualized has significant implications for the validity of test scores. There was no evidence to support for grouping men according to two groups, those who are pedophilic and those who are not pedophilic. There was some evidence to support a continuous model of pedophilic interest, however this operationalization did not predict above static risk. There was more support for operationalizing pedophilic interests trichotomously, however, within this model, only the two-level model performed consistently across analyses. The results also suggest that pedophilic interests, when operationalized trichotomously, provide unique information beyond considerations of static risk. This is an important consideration, as few psychologically meaningful risk factors have been examined in terms of their ability to add unique information to static risk. A notable limitation in the current study that limits the strength of this interpretation is that the sample size for the preferentially pedophilic men was small (i.e., $n = 14$).

These latent structure results are consistent with the past body of research using different methods of operationalizing pedophilic interest. One limitation of the existent literature is that past studies provide a partial or indirect test of different latent structural models of pedophilic interest. That is, much of the past research has chosen one or two methods of operationalizing pedophilic interests and have provided various validity tests of that operationalization. For instance, Moulden et al. (2009) examined whether a dichotomous operationalization predicted

sexual recidivism; while Wilson et al. (2011) examined both dichotomous and trichotomous operationalizations. This is not a limitation of these studies per se, as the authors were not necessarily attempting to test latent structural models and taxometric results were unavailable to much of the past research into the validity of measures of pedophilic interest. However, the present analyses provide a unique contribution to our understanding of which latent structures is a more optimal operationalization of pedophilic interest. The present results expand the previous findings and provide further and strong evidence that considering preferentiality of pedophilic interest is a valid indicator of sexual recidivism risk.

5.4.1 Limitations

The present study did not include the revised version of the SSPI, the SSPI-2. The SSPI-2 has been found to have improved psychometric properties and to correlate with phallometric measures of pedophilic interest to a somewhat stronger degree than the original scale. We would anticipate that the inclusion of an item regarding online offending would improve the relationship between the SSPI and the other measures used in this study, given the evidence that men who access online child sexual exploitation materials tend to be more pedophilic (Seto, Cantor, & Blanchard, 2006). As a result, the validity of the SSPI presented in this study may be somewhat attenuated from the validity estimates the SSPI-2 may produce in a similar sample.

A further limitation is that pedophilia diagnosis was not available in this sample, negating our ability to examine the sensitivity and specificity of the VRS-SO Sexual Deviance factor for diagnosis. As well, we were unable to assess the correlation between VSO-SO assessed sexual deviance and diagnosis, which would have contributed novel findings to existing research showing a correlation between the VRS-SO static, dynamic, and total score and exclusive

pedophilia diagnosis (Eher et al., 2015). This set of analyses would provide further evidence for the VRS-SO Sexual Deviance factor as capturing pedophilic interest.

5.4.2 Conclusion

The three measures of pedophilic interests examined had varying degrees of validity support. Phallometric measures and the VRS-SO Sexual Deviance factor score were shown to have consistent convergent validity evidence and relatively stable associations with sexual recidivism. The SSPI had less validity support, with the most notable being a lack of association between SSPI scores and sexual recidivism.

Pedophilic interest is a robust predictor of sexual recidivism (McPhail et al., 2017), while recent research shows that there is not always a stable relationship (Stephens et al., 2017). One potential explanation for the variation in results is that different studies use different methods to operationalize the latent structure of pedophilic interest. The present study examined the validity of different operationalizations and found the most support for a trichotomous latent structure that grouped men into those with preferential pedophilic interest and those with non-preferential pedophilic interest or no pedophilic interest. These results are consistent with past research that has considered preferentiality or exclusivity of pedophilic interests. Future research should replicate these validity tests and tests of latent structure in measures of pedophilic interest. Larger sample sizes are required for such research, especially since pedophilic interest is more generally a risk factor within samples of men with offences against children and the base rate of preferential pedophilic interests within such samples is low.

Table 5-1.

Correlations among measures of pedophilic interest.

	PPG – Child	PPG – Female	PPG – Male	SSPI	VRS-SO Sexual Deviance pre-treatment	VRS-SO Sexual Deviance post-treatment
PPG – Child	--	.93***	.78***	.38[a]***	.39***	.42***
		(261)	(261)	(74)	(208)	(208)
PPG – Female	.90***	--	.65[a]***	.25[a]*	.33***	.35***
	(91)		(261)	(74)	(208)	(208)
PPG – Male	.82***	.54[a]***	--	.46[a]***	.35***	.37***
	(91)	(91)		(74)	(208)	(208)
SSPI	.24	.09	.41[a]*	--	.43***	.39***
	(48)	(48)	(48)		(79)	(79)
VRS-SO Sexual Deviance – Pretreatment	.35**	.27*	.35**	.49***	--	.98***
	(76)	(76)	(76)	(50)		(225)
VRS-SO Sexual Deviance – Posttreatment	.39***	.31**	.39**	.45**	.98***	--
	(76)	(76)	(76)	(50)	(83)	

Note. Correlations below the diagonal are for the sexual offender against children sample. Correlations above the diagonal are for the entire sexual offender sample. Sample sizes are in parentheses below the correlation coefficient. PPG = Phallometric testing. SOC = Sexual offenses against children. VRS-SO = Violence Risk Scale-Sexual Offense Version.

[a]Spearman's rho coefficient.

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

Table 5-2.

Predictive accuracy of measures of pedophilic interest for sexual recidivism.

Measure	Aggregate sample (n = 250)		Sexual offenses against children (n = 89)	
	AUC	[95% CI]	AUC	[95% CI]
Pedophilic stimuli – Child	.61**	.54, .69	.71**	.59, .83
Pedophilic stimuli – Female	.60**	.53, .68	.69**	.57, .82
Pedophilic stimuli – Male	.57	.49, .65	.65*	.53, .77
SSPI	.56[a]	.42, .69	.63[b]	.47, .79
VRS–SO Sexual Deviance Factor				
Pre-treatment	.60*[c]	.52, .68	.71**[d]	.59, .83
Post-treatment	.62**[c]	.54, .71	.69**[d]	.57, .82

Note. VRS-SO = Violence Risk Scale-Sexual Offense Version.

[a] n = 79. [b] n = 50. [c] n = 225. [d] n = 85.

* $p < .05$. ** $p < .01$.

Table 5-3.

Predictive accuracy of measures of pedophilic interest for sexual recidivism.

Predictor	*n*	*B*	*SE*	Wald	e^B	95% CI for e^B
PPG – Child[a]	245	.58	.23	6.54*	1.79	1.15, 2.79
SOC only	87	.04	.01	11.22**	1.04	1.02, 1.06
PPG – Female child	245	.10	.01	5.66*	1.01	1.00, 1.02
SOC only	87	.02	.01	12.37***	1.02	1.01, 1.04
PPG – Male child	245	.01	.01	2.57	1.01	0.99, 1.03
SOC only	87	.02	.01	6.73**	1.02	1.01, 1.04
SSPI	245	.06	.13	0.18	1.06	0.81, 1.37
SOC only	50	.27	.19	1.98	1.31	0.90, 1.91
VRS-SO – Pretreatment	225	.07	.03	5.54*	1.07	1.01, 1.14
SOC only	83	.15	.06	6.17*	1.16	1.03, 1.31
VRS-SO – Posttreatment	225	.10	.03	8.34**	1.10	1.03, 1.17
SOC only	83	.16	.06	6.96**	1.18	1.04, 1.33

Note. PPG = Phallometric testing. SOC = Sexual offenses against children. VRS-SO = Violence Risk Scale-Sexual Offense Version.

[a]Log transformed variable used.

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

Table 5-4.

Incremental contribution for measures of pedophilic interest to the prediction of sexual recidivism.

Predictor	χ^2 Change	*B*	*SE*	Wald	e^B	95% CI for e^B
PPG – Child[ab]						
Step 1[c]						
Static-99R		.31	.06	27.21***	1.36	1.21, 1.53
Step 2						
Static-99R		.29	.06	25.02***	1.35	1.20, 1.52
PPG – Child	1.72	.01	.01	1.85	1.01	0.99, 1.03
SOC only						
Step 1[c]						
Static-99R		.37	.09	16.93***	1.45	1.22, 1.73
Step 2						
Static-99R		.32	.10	1.98***	1.37	1.14, 1.65
PPG – Child	2.40	.02	.01	2.62	1.02	0.99, 1.05
PPG – Female child[b]						
Static-99R		.30	.06	24.78***	1.35	1.20, 1.51
PPG – Female child	1.95	.01	.004	2.11	1.01	0.99, 1.02
SOC only						
Static-99R		.30	.10	10.27**	1.36	1.13, 1.63
PPG – Female child	2.89	.01	.01	3.13	1.01	0.99, 1.03
PPG – Male child[b]						
SOC only						
Static-99R		.36	.09	14.72***	1.43	1.19, 1.72
PPG – Male child	2.29	.02	.01	2.63	1.02	0.99, 1.03
VRS-SO – Pretreatment						
Static-99R		.25	.06	17.25***	1.29	1.14, 1.45
VRS-SO – Pretreatment[d]	3.33†	.06	.03	3.37	1.06	0.99, 1.12
SOC only						
Static-99R		.21	.09	5.13*	1.24	1.03, 1.49
VRS-SO – Pretreatment[e]	2.26	.10	.07	2.15	1.10	0.97, 1.25
VRS-SO – Posttreatment						
Static-99R		.25	.06	16.60***	1.28	1.14, 1.44
VRS-SO – Posttreatment[d]	5.26*	.08	.03	5.39*	1.08	1.01, 1.15
SOC only						
Static-99R		.21	.09	5.29*	1.24	1.03, 1.48
VRS-SO – Posttreatment[e]	2.99	.12	.07	2.95	1.12	0.98, 1.28

Note. PPG = Phallometric testing. SOC = Sexual offenses against children. VRS-SO = Violence Risk Scale-Sexual Offense Version.

[a]Log transformed variable used. [b]n = 241. [c]The results from the first step in the analyses are the same for each model and are not repeated in this table. [d]n = 225. [e]n = 83.

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

Table 5-5.

Prediction of sexual recidivism using survival analysis across three models of pedophilic interest.

Predictor	*n*	*B*	*SE*	Wald	e^B	95% CI for e^B
Pedophilic Interest – Overall						
Continuous model[a]	245	.58	.23	6.54*	1.79	1.15, 2.79
SOC only	87	.04	.01	11.22**	1.04	1.02, 1.06
Dichotomous model	245	.10	.28	0.14	1.11	0.65, 1.91
SOC only	87	.60	.42	2.05	1.83	0.80, 4.17
Trichotomous model (3-levels)	245	.19	.20	0.98	1.21	0.83, 1.78
SOC only	87	.75	.29	6.57*	2.12	1.19, 3.78
Trichotomous model (2-levels)	245	.69	.40	2.99	1.99	0.91, 4.34
SOC only	87	1.79	.51	12.16***	5.99	2.19, 16.40
Pedophilic Interest – Female orientation						
Continuous model	245	.01	.004	5.66*	1.01	1.00, 1.02
SOC only	87	.02	.007	12.37***	1.02	1.01, 1.04
Dichotomous model	245	.15	.27	0.30	1.16	0.69, 1.95
SOC only	87	.50	.43	1.38	1.65	0.72, 3.82
Trichotomous model (3-levels)	245	.16	.21	0.60	1.18	0.78, 1.78
SOC only	87	.64	.33	3.77	1.89	0.99, 3.61
Trichotomous model (2-levels)	245	.47	.51	0.84	1.60	0.58, 4.39
SOC only	87	1.67	.62	7.18**	5.31	1.57, 17.99
Pedophilic Interest – Male orientation						
Continuous model	245	0.01	.01	2.569	1.01	0.99, 1.03
SOC only	87	0.02	.01	6.73**	1.02	1.01, 1.04
Dichotomous model	245	0.34	.26	1.68	1.41	0.84, 2.35
SOC only	87	0.43	.42	1.05	1.54	0.68, 3.51
Trichotomous model (3-levels)	245	0.31	.19	2.63	1.36	0.94, 1.98
SOC only	87	0.53	.27	3.91*	1.71	1.01, 2.90
Trichotomous model (2-levels)	245	0.67	.43	2.48	1.96	0.85, 4.51
SOC only	87	1.29	.48	7.35**	3.64	1.43, 9.27

Note. PPG = Phallometric testing. SOC = Sexual offenses against children. VRS-SO = Violence Risk Scale-Sexual Offense Version.

[a]Log transformed variable used.

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

Table 5-6.

Incremental contribution for three models of pedophilic interest of the prediction of sexual recidivism using survival analysis.

Predictor	Change	*B*	*SE*	Wald	e^B	95% CI for e^B
Pedophilic Interest – Overall Child						
Continuous model[abc]						
Static-99R		.29	.06	25.02***	1.34	1.20, 1.51
Pedophilic interest	1.72	.41	.23	1.85	1.51	0.97, 2.36
SOC only						
Static-99R		.32	.10	1089***	1.37	1.14, 1.65
Pedophilic interest	2.40	.02	.01	2.62	1.02	0.99, 1.05
Trichotomous model (3-levels)						
SOC only						
Static-99R		.34	.01	12.44***	1.41	1.16, 1.70
Pedophilic interest	0.75	.29	.78	0.78	1.34	0.70, 2.55
Trichotomous model (2-levels)						
SOC only						
Static-99R		.34	.09	13.10***	1.40	1.17, 1.69
Pedophilic interest	3.88*	1.16	.53	4.70*	3.18	1.12, 9.07
Pedophilic Interest – Female Orientation						
Continuous model						
Static-99R		.30	.06	24.78***	1.35	1.20, 1.51
Pedophilic interest	1.95	.01	.004	2.11	1.01	0.99, 1.02
SOC only						
Static-99R		.30	.10	10.27**	1.36	1.13, 1.63
Pedophilic interest	2.90	.01	.01	3.13	1.01	0.99, 1.03
Trichotomous model (2-levels)						
SOC only						
Static-99R		0.36	.09	15.67***	1.44	1.20, 1.72
Pedophilic interest	3.12	1.30	0.64	4.18*	3.67	1.06, 12.79
Pedophilic Interest – Male Orientation						
Continuous model						
Static-99R		0.36	.09	14.72***	1.43	1.19, 1.72
Pedophilic interest	2.29	0.02	.01	2.63	1.02	0.99, 1.03
Trichotomous model (3-levels)						
SOC only						
Static-99R		0.36	.09	14.59***	1.43	1.19, 1.72
Pedophilic interest	0.32	0.17	.29	0.33	1.18	0.67, 2.08
Trichotomous model (2-levels)						
SOC only						
Static-99R		0.36	.09	14.35***	1.43	1.19, 1.72
Pedophilic interest	2.32	0.79	.48	2.64	2.19	0.85, 5.66

Note. PPG = Phallometric testing. SOC = Sexual offenses against children. VRS-SO = Violence Risk Scale-Sexual Offense Version.

[a]Log transformed variable. [b]These results are the same as the PPG – Child Cox regression results presented in Table 4-4. [c]The results from the first step in the analyses are the same for each model and are not repeated in this table.
* $p < .05$. ** $p < .01$. *** $p < .001$.

CHAPTER 6

GENERAL DISCUSSION

This dissertation aimed to examine three issues in how pedophilic interest is conceptualized, treated, and assessed. The taxometric study provided interesting empirical evidence regarding how best to conceptualize pedophilic interest and these results meaningfully informed the latter two studies in the dissertation. The meta-analytic review found that, in general, interventions were associated with reductions in pedohebephilic interests. The third study found multiple measures of pedophilic interest converge and are predictive of sexual recidivism. The latter two studies of this dissertation also provide meaningful data on the validity of different latent structural models of pedophilic interest.

The present research found that the latent structure of pedophilic interest in men may be characterized as forming three qualitatively distinct classes, which were characterized as non-pedophilic, non-preferentially pedophilic, and preferentially pedophilic. A three-class structure affirms theoretical expectations regarding the latent structure of pedophilic interest (Hanson, 2010; Seto, 2017), while extending these in an interesting direction. Namely, that preferentially pedophilic men may be qualitatively different from other men, even those with non-preferential pedophilic interest. A three-class structure seems to fit with pre-existing clinical operationalizations of pedophilic interest, specifically Pedophilic Disorder in the DSM-5 (APA, 2013).

6.1 Taxometric Replication Studies are Needed

A general tenant of the taxometric approach to identifying taxonicity or dimensionality is consistency testing. That is, disparate statistical procedures are used in order to identify whether the same result is produced across the procedures. Meehl (1992) provided consistency testing as

a means for taxometric methods to provide a "Popperian risk of strong discorroboration" (p. 1). A scientific ideal is at work within the conceptual assumptions of taxometric methods: In order to believe that a certain latent structure characterizes a construct, the latent structure has to survive severe tests that run the risk of not corroborating that the latent structure characterizes the construct. Without exposing potential latent structural conclusions to the risk of discorroboration, latent structural conclusions are on shaky empirical ground, and should be subjected to heavy skepticism.

The interpretive approach taken in Chapter 2 that resulted in pursuit of trichotomous latent structure was indeed liberal. The justification provided here for taking a liberal approach is that without an eye to alternatives to explain patterns in the data, making a kind of Type II error was deemed an unacceptable risk at this early stage of research into the latent structure of pedophilic interest.[19] This error may cause us to miss interesting alternatives by blinding us to focus solely on two possible outcomes: dichotomous or dimensional latent structure. The obvious drawback to reducing Type II error is that this approach increases the risk of Type I error (i.e., concluding some latent structure exists when it in fact does not). The results of the present taxometric analysis may also be explained by dimensional structure to pedophilic interest, with the so-called preferential pedophilic taxon being the upper end of this dimension. This may have been the case because the present study was simply not able to detect dimensionality. This returns us back to the necessity of exposing latent structural results to the risk of strong discorroboration, via

[19] An interesting posthoc justification, "Open-minded empirical exploration of taxonic possibilities will, I am sure, lead to some interesting surprises. *Example:* We do not usually consider *severity* of a specific organic disease as taxonic, rather we view severity as involving one or more quantitative components within a taxon. But Hoagland (1947), in a fascinating study, showed [via plots of EEG frequency] early, intermediate, and advanced paresis as discrete categories." (Meehl, 1992, p. 129; emphasis in original). Such open-mindedness is a hallmark of scientific exploration, especially in relatively uncharted and underdeveloped areas of research, such as basic science into latent structure of human sexuality.

replication and consistency testing. That is, interesting alternatives must be given a fair chance to fail.

Reasonable scientific skepticism requires replication studies to illuminate which latent structural model(s) receive support and which do not survive repeated tests. It is still possible that pedophilic interest represents a dimension of increasing severity in terms of interest in and preference for children. Replication studies may take several forms. Such studies may use statistical methods other than, or in conjunction with taxometric analysis to model the latent structure of pedophilic interest. For instance, a useful adjunct to taxometric analysis is latent profile analysis (see Schmidt et al., 2013). Latent profile analysis provides indices of fit for models with dimension, two-class, three-class, etc. structures, which would allow for additional evaluation of how well different models fit the observed data (Borsboom et al., 2016).

One potential limitation to the existing research base is that each study conducted to date has used different sets or combinations of measures. An important next step is to conduct a latent structural study using the same set of measures as a previous study but in a unique sample. The battery of measures used by Schmidt et al. (2013) is a strong candidate for replication, as the measures used have a reasonable body of validity evidence available (Banse, Schmidt, & Clarbour, 2010; Schmidt, Gykiere, Vanhoeck, Mann, & Banse, 2014) and includes multimodal assessment of pedophilic interest. A replication study using the same measures as Schmidt et al. (2013) with a large sample of men would represent a strong replication test of the existing latent structural models.

The necessity of a large sample size arises from the very low base rate of a putative preferential pedophilic taxon, especially in a sample comprised of men who have no history of sexual offending. One methodological solution to the problem of a low base rate in a putative

preferentially pedophilic taxon is to conduct replication research in a two-step manner. The first step would be to conduct latent structural analyses in a large sample of men drawn from the general population. If a small pedophilic taxon emerges in these analyses, this would be reasonably strong evidence for the existence of a pedophilic taxon. The second step would be to provide the same set of measures to a sample of self-identifying pedophilic men, which represent a sample of men within a pedophilic taxon identified in the first step. Latent structure analyses within this sample of self-identifying pedophilic men would provide a test of the latent structure of the pedophilic taxon identified in the first step. The results of the second step may provide further understanding of whether a pedophilic taxon itself is dimensional or categorical.

A few findings in Chapter 2 provide further evidence of the need for replication of the present taxometric results. The elevated rate of sexual compulsivity in the two pedophilic taxa makes the interpretive picture of what the taxa consist less clear. The middle taxon may have emerged because these men experience high sex drive and therefore show elevated rates of sexual arousal to both child and adult erotic stimuli. This interpretation suggests that upon replication, measures that are not sensitive to sex drive (e.g., rating or viewing time measures) would result in a dichotomous latent structure. However, given that hypersexuality appears to be dimensional (Graham, Walters, Harris, & Knight, 2015; Kingston et al., 2018), the results presented in Chapter 3 provide some limited evidence against this conclusion. Future research will need to consider how sex drive, hypersexuality, sexual compulsivity, and responding on measures of pedophilic interest may be related and the effects of this on taxometric results.

A further line of structural research that is needed would analyze measures of pedophilic interest to factor analyses. The taxometric study conducted here and the replication studies outlined above will likely provide little information about to what degree pedophilic and

teleiophilic interests co-occur in men. Factor analytic studies, on the other hand, will allow for a better understanding of whether pedophilic, hebephilic, and teleiophilic interests are better conceptualized as relatively distinct or co-varying sexual interest patterns. These studies may face interesting problems that will require interesting methodological and analytic solutions. For instance, covariance between pedophilic interest and sexual compulsivity has been replicated in several studies, and it may be reasonable to expect the covariance between pedophilic and teleiophilic interests will be saturated with sex drive variance. Because sexual interest is also structured according to the sex of people one is attracted to, future factor analytic studies will need to consider both sex and age. What this may require is an analytic approach that tests the fit of multiple factor analytic models, including models with hierarchical structure to capture strength of sex drive as an over-arching construct.

6.2 Possible Etiological Pathways to Two Pedophilic Taxa

This is perhaps a reasonable place in the program of taxometric research to ask the question: if there are indeed two pedophilic taxa, is there reason to believe there are different etiological pathways that explain why an individual is a member of one pedophilic taxon while another individual is a member of the other taxon? Examination of this question may profit by starting from a general perspective on how etiological causes are related to taxonicity.

The strongest conceptualization of causality is specific etiology, which can take the following forms: an etiological cause is dichotomous and is necessary and sufficient for the taxonic characteristic to be present; an etiological cause is dichotomous and is indispensable for the presence of the taxonic characteristic, but is not sufficient and the presence of other etiological (and potentially nonspecific) causes is required; or an etiological cause is quantitative, and the probability of the presence of the taxonic characteristic is zero prior to some threshold in

the etiological cause (i.e., specific threshold etiology; Meehl, 1977, 1992). Weaker, nonspecific etiological pathways are also possible for the presence of a taxonic characteristic, which include: multiple unique and non-specific factors are etiological causes and can be present to varying degrees between individuals, but some etiological causes have greater influence on the presence of the taxonic characteristic; the existence of a variety of unique and non-specific etiological factors and some yet-to-be-determined combination of the presence and absence of these factors leads to the presence of the taxonic characteristic; or, quantitative etiological factors that, instead of having a threshold dividing taxon from non-taxon members, are stepped in terms of the probability of the presence of the taxonic characteristic (Meehl, 1977). An orthogonal concept to elaborate and combine with the cases of specific and non-specific causality is divergent causality. In divergent causality, variations in causal factors at an early stage accumulate and cause changes within individuals that increase across time (Meehl, 1992). As can be seen, there are multitudinous possible ways of conceptualizing etiology of taxonic characteristics, and the following discussion does not aim to be exhaustive of these possibilities.

In considering etiological pathways of pedophilia, if we are to believe there are two discrete pedophilic taxa, it will be important to consider whether there are unique or non-specific causes that lead to the two putative pedophilic taxa. In the context of the etiology of pedophilic interest, it may be the case that: there is a unique causal pathway to the preferential pedophilic taxon and there is a separate unique causal pathway to the non-preferentially pedophilic taxon; there are multiple causal factors that are specific to the presence of preferential pedophilic interest; there are multiple causal factors that are specific to the presence of non-preferential pedophilic interest; there are multiple causal factors specific to the development of pedophilic interest, in general, and preferentially pedophilic men experience more of these factors; or, there

are some range of specific and non-specific causal factors for the developmental of pedophilic interest, in general, and preferentially pedophilic men experience more of these factors. Taking divergent causality into consideration in relation to these possibilities, the experience of causal factors at different developmental periods (i.e., in utero or very early in life vs. late childhood or early adolescence) might differentially contribute to the presence of preferential versus non-preferential pedophilic interest.

An important limitation to identify before discussing answers to the question of etiology of pedophilic interest is that the research into causal mechanisms for pedophilic interest is underdeveloped. At present, we may propose hypotheses about etiological differences between two pedophilic taxa, but evidence, in the strong sense that we can point to a set of well-established etiological factors, is currently lacking.

A first possible etiological pathway that differs between the two pedophilic taxa comes from the biogenic hypothesis. The biogenic hypothesis suggests that men who are preferentially pedophilic are more likely to experience biological and environmental factors that contribute to the development of pedophilic interest from very early in life, perhaps beginning *in utero* (King, 2010; McPhail & Cantor, 2015). This biogenic, or neurodevelopmental perturbations, hypothesis makes a set of hypotheses about the factors that may increase vulnerability to developing pedophilic interest. These men may have: higher rates of neurodevelopmental disorders, which tend to co-occur (Rommelse, Franke, Geurts, Hartman, Buitelaar, 2010; Siminoff, Pickles, Charman, Chandler, Loucas, & Bard, 2008; Zauche, Darcy Mahoney, & Higgins, 2017); cognitive deficits; poor scholastic aptitude in childhood; higher rates of fraternal older brothers; an earlier onset of attraction patterns suggestive of pedophilic interest; and anatomical signs of in utero difficulty, such as low birth weight, shorter stature, more non-righthandedness, and more

minor physical anomalies (Brankely, 2019; Dyshniku, Murray, Fazio, Lykins, & Cantor, 2015; Fazio, Lykins, & Cantor, 2014; King, 2010; McPhail & Cantor, 2015). It should be noticed that none of these causal factors are specific to pedophilic interest but are features of non-heterosexuality and other disorders as well (e.g., schizophrenia; Owen, O'Donovan, Thapar, & Craddock, 2011).

In more general research on sexual orientation, homosexual men tend to show higher rates of biogenic markers, such as shorter stature or a greater number of older maternal brothers (Blanchard, 2018; Skorska & Bogaert, 2017), than heterosexual men. Further research has found homosexual men to be shorted than heterosexual men, but bisexual men's height is between the two groups and not different from heterosexual men (Skorska & Bogaert, 2016). We may extend the findings with sexual orientation and expect that exclusivity or preferentiality of age orientations directed towards the non-normative targets (i.e., prepubescent children) will be associated with more markers of biogenic etiology. The result of this extension would be that preferentially and exclusively pedophilic men are predicted to be more likely than non-preferentially pedophilic and non-pedophilic men to: have older brothers by the same mother, be diagnosed with a neurodevelopmental disorder, be shorter, have more minor physical anomalies, be non-right handed, and have an earlier age of onset of attractions that follow a pedophilic pattern (i.e., the targets of their first attractions in late childhood are much younger than themselves). There are likely additional hypotheses that can be developed to expand biogenic etiological explanations, and this short list should not be considered exhaustive.

This general hypothesis can be made more specific in several ways. It may be that preferentially pedophilic men will show elevated rates of these putative etiological factors, while there will be no differences between non-preferentially pedophilic and non-pedophilic men. This

would a biogenic pathway to preferential pedophilic interest, but not pedophilic interest in general. Alternatively, it may be that preferentially pedophilic men show the highest rates of these putative etiological factors, while non-preferentially pedophilic men show lower rates than this group but higher rates than non-pedophilic men. This increase may be stepped in that non-preferentially pedophilic men have experienced a rate above some threshold and preferential men have experienced rates above a further threshold. This would suggest a common biogenic pathway to pedophilic interest, in general, with preferential men experiencing more of the putative etiological factors. Although other possibilities remain, such as non-preferentially pedophilic men may experience higher levels of biogenic factors, however this seems counter intuitive and unlikely to be supported in future research.

Research has found that these neurodevelopmental vulnerabilities represent a discrete taxon (King, 2010). This research suggests, that within sexual offenders at least, vulnerabilities that are thought to be putative causes of pedophilic interest are taxonic. Of note, individuals who were identified as belonging to this neurodevelopmental vulnerability taxon were characterized as being more likely to have victimized a prepubescent child and less likely to have been married. These two findings suggest that there is an association between these neurodevelopmental vulnerabilities and proxies for pedophilic interest, and potentially preferential pedophilic interest because we might hypothesize that one reason men, who sexually offend against children, do not marry is that they are not sexually attracted to adults.

Recently acquired data may provide some insight into the presence biogenic factors in pedophilic taxa. In a sample recruited online, exclusive pedophilic men (n = 110), non-exclusive pedophilic men (n = 107), and teleiophilic men (n = 120) self-reported past diagnoses of neurodevelopmental disorders, adverse childhood experiences, and the age of the target of their

first sexual attractions (McPhail & Stephens, 2019). Conforming to the above-mentioned hypotheses, exclusively pedophilic men were more likely to report a history of neurodevelopmental disorders (23% vs. 15% in non-exclusively pedophilic men and 6.7% in teleiophilic men). Further, exclusively pedophilic men reported that the objects of their first attractions were much younger (average age = 8.1 years) compared to non-exclusively pedophilic (average age = 10.2 years) and teleiophilic men (average age = 13.0 years). These findings lend some support for a shared, stepped pathway of biogenic factors, with preferentially pedophilic men experiencing more of these risk factors.

Another recent taxometric analysis found that pedophilic interest has taxonic latent structure (Brankely, 2019). That author did not examine the latent structure of the pedophilic taxon; however, the neurodevelopmental features of the pedophilic taxon and complement class were examined. Across two separate samples of sexual offenders, significant differences were found between pedophilic taxon and complement class members in physical height, late maturation, intelligence, mental retardation, and attending special classes; though these last three indicators are likely highly conflated. It is important to note that the majority of the neurodevelopmental perturbations examined in that study were similar in the two taxa and the differences that were found were small in magnitude. These results are hard to interpret, as it may be the case that there generally is no difference between pedophilic men and non-pedophilic men on these biogenic factors, or the relatively small differentiation occurred because that author did not consider preferentiality of pedophilic interest.

A second possible pathway is that pedophilic interest results from psycho-sociogenic factors. A main hypothesis within psycho-sociogenic etiology is that childhood adversity is in some way linked to developing pedophilic interest. Another interesting hypothesis from the

psycho-sociogenic perspective is whether childhood sexual experiences are associated with adulthood pedophilic interest (Santilla et al., 2010). Putative psycho-sociogenic etiological factors, then, may include experiencing adversity in middle childhood or adolescence and sexual experiences with other children in middle childhood (i.e., under the age of 10 to 12). Given the anticipated developmental periods of these experiences, we may further expect that psycho-sociogenic causation will be associated with objects of first sexual attractions to be same age peers (McPhail, 2018).

This general psycho-sociogenic hypothesis can also be made more specific in several ways. As a first step in identifying these hypotheses, it seems more plausible that men who experience sexual attractions to both children and adults are more likely to experience psycho-sociogenic factors. This is because these factors might be better conceptualized as experiences later in childhood that interfere with, but do not fundamentally alter, typically developing sexual attractions to similar age peers in adolescence and into adulthood. This assumption is derived from divergent causality in that factors accumulated later in life lead to less divergence from typical development (i.e., non-preferentiality is a lesser divergence from teleiophilia than preferential pedophilia). Given this, we might expect that: non-preferentially pedophilic men experience more psycho-sociogenic factors, while preferentially pedophilic and non-pedophilic men will report relatively similar rates of these putative etiological factors. Support for this specific hypothesis would suggest that there is a psycho-sociogenic causal pathway to non-preferential pedophilic interest. That is, men who experience sexual attraction to children, but also retain the potential and ability to be sexual attracted to adults, experience higher rates of psycho-sociogenic factors. A stepped pattern of rates may also be plausible, such that non-preferential men experience rates of psycho-sociogenic factors higher than preferentially

pedophilic and non-pedophilic men, while preferentially pedophilic men experience higher rates than non-pedophilic men. Alternatively, in combination with the biogenic hypothesis, preferentially and non-preferentially pedophilic men may exhibit the same rates of psycho-sociogenic factors that are elevated compared to non-pedophilic men. However, preferentially pedophilic men are the pedophilic men who also experience biogenic factors.

There is empirical support for an association between childhood adversity and pedophilic interest in adulthood. In two large samples, an association was found between sexual abuse and sexual interest in children in adulthood (Alanko et al., 2017; Santilla et al., 2010). This association was also found for emotional neglect (Alanko et al., 2017). In a separate analysis of the Santilla et al. (2010) sample, an association between sexual experiences with other children and sexual interest in children under the age of 16 in adulthood was found (Santilla et al., 2010). Importantly, that study examined the influence of genetic factors on sexual interest in children and those researchers concluded that sexual interactions with other children represents true environmental causation. Recent data also suggest that rates of sexual abuse in childhood are higher in non-exclusively pedophilic men (31.8%) compared to exclusively pedophilic (19.6%) and teleiophilic men (14.2%; McPhail & Stephens, 2019). One main caveat is that the research designs used were cross-sectional, which cannot be used to interpret causation of these putative risk factors.

Taken together, these theoretical hypothesis and empirical research suggests that there is an array of combinations for causal pathways for the development of pedophilic interest. Given the present state of knowledge regarding both latent structure in and causal risk factors for pedophilia, there are meaningful directions for future research. Such research may aim to

identify how putative causal factors are distributed in two pedophilic taxa or whether there are causal factors that are specific to pedophilic interest.

6.3 Effects of Treatment on Pedohebephilic Interests

There are a few noteworthy findings in the meta-analytic review of interventions for pedohebephilic interests. One was that most treatments were associated with decreases in sexual arousal to children. For those treatments that were effective, there was a kind of dodo bird effect: No treatment modality was clearly superior to others in terms of the magnitude of treatment effects. This dodo bird finding aligns with much of the available evidence regarding the efficacy and effectiveness of psychotherapy for other mental health conditions (e.g., depression, Furukawa et al., 2017). A clinical implication of this set of results is that, at present, clinicians can offer their clients a suite of treatment options and inform clients that each intervention is anticipated to have equal effectiveness. The importance of this is that some clients will not want to undergo certain forms of treatment (e.g., pharmacological intervention due to side-effects or aversive interventions due to discomfort associated with these procedures). Interestingly, studies combining pharmacological and behavioural interventions are currently lacking; there is evidence to suggest that combining psychotherapy and pharmacotherapy can improve outcomes (Kamenov, Twomey, Cabello, Prina, & Ayuso-Mateos, 2017).

Other findings of interest were that treating pedohebephilic interest within the context of a comprehensive treatment program resulted in little positive change. This finding is as puzzling as it is interesting. We have outlined a few potential explanations above, but it is worth noting that most men who participate in sexual offense treatment do so within the context of comprehensive programs (McGrath et al., 2010). This may suggest that in current practice, the majority of men

receive services in such a manner that may not result in a meaningful change in their ability to manage their arousal to children.

We pursued the implications of latent structure within the treatment effectiveness meta-analysis. Grouping samples according to trichotomous latent model significantly moderated the effect of treatment: highly pedophilic men showed the largest amounts of treatment gain. An analogous way to conceptualize these findings is that those with more severe baseline symptoms of pedophilic interest demonstrated the greatest change over the course of treatment. This finding is significant and of clinical importance. If this result holds to be true in future treatment efficacy studies, this will suggest that those men who experience the most intense pedophilic interests are able to manage their sexual interest in children. This result would be a strong indication that clinical pessimism regarding the treatability, or manageability of strong pedophilic interest is not warranted.

Such a state of affairs has significant clinical implications regarding what clinicians are able to tell their clients in terms of treatment expectancies and the management of men convicted of sexual offences. Replication would suggest those at most risk could demonstrate their ability to manage their arousal in the community and could mean less time incarcerated. The findings that psychotherapy can help men with higher levels of pedophilic interest change joins a growing body of research suggesting that psychotherapeutic approaches can benefit even those with the most severe symptomatology (Furukawa et al., 2017). In the case of interventions for pedohebephilic interests in men with sexual offense histories, this may be an especially positive development, as the side-effects of long-term anti-androgen use can be debilitating and life threatening (Nota et al., 2019; Turner & Briken, 2018). However, it is important to note that at

present, there is little to no information regarding adverse reactions clients have to behavioural treatments for pedophilic interest.

The implications of this moderating effect of symptom severity are large. For instance, this result aligns with the well-known principles of effective offender rehabilitation: services are provided for specific biopsychosocial risk factors that are demonstrably present for clients and the intensity of services are titrated to the intensity of the need (Bonta & Andrews, 2017). Examination of the size of the samples in the meta-analysis suggests that a small minority of men were highly pedophilic and a minority were moderately pedophilic. However, across the studies, a large number of men received these specialized interventions for managing arousal to children who may not have experienced a need in this domain. To the extent that this is true in the programs evaluated by these studies, and the extent that it continues to happen in clinical practice, this represents a waste of treatment services. Given the potential adverse reactions to aversive interventions, clients are subjected to potentially harmful and distressing treatments that they do not require. The results of the meta-analysis can be used to guide the administration of clinical practice to provide treatment only to those with moderate or high levels of pedohebephilic interest.

While meta-analytic studies tend to be reviews of existing data, we were able to contribute some novel findings to the existent body of research. By constructing natural history benchmarks, we identified that the majority of interventions were associated with changes that were clinically and statistically greater than the amount of change due to natural history processes. In addition, men with a high level of pedohebephilic interest showed more change that expected due to these natural processes. The importance of this set of analyses is that it increases confidence in the findings by allowing a conclusion that treatment change is likely not (solely) a

result from natural processes. The normative comparisons also extended the understanding of treatment change in pedohebephilic interest. Approximately half of the analyses supported the conclusion that men who underwent treatment showed levels of arousal to children that were similar to men with no history of sexual offending against children. While this is not unequivocal evidence, given the necessity to piece normative comparisons together from different studies and the lack of consistency in the results, it provides some support for the conclusion that offending men do show similar levels of arousal compared to non-offending men.

These meta-analytic results add to an expanding body of literature that suggests treatment works and that changes on dynamic risk factors leads to meaningful reductions in sexual recidivism. A recent meta-analysis suggests that treatments that target arousal control lead to greater reductions in sexual recidivism (Gannon et al., 2018). The present meta-analytic results buttress this finding by indicating that this reduction in sexual recidivism may indeed be due to the effectiveness of interventions for pedohebephilic interests. While the present review did not examine whether treatment change was associated with reductions in sexual recidivism, research has found that changes on dynamic risk factors, which pedophilic interests may be considered to be, over the course of treatment predict reductions in sexual recidivism (Olver, Kingston, Nicholaichuk, & Wong, 2014). Other meta-analytic reviews find that general change on dynamic sexual violence risk assessment measures is also associated with decreased sexual recidivism (Van den Berg et al., 2018).

The meta-analytic results suggest a few lines of future research. A first step will involve conducting randomized control trials of existing or newly developed interventions for pedohebephilic interest. While the existent body of research provides reason for optimism that treatments are the cause of the decrease in pedohebephilic arousal, the designs used in the studies

included in the meta-analysis rule out a conclusion that this is indeed the cause of the change. The results of the meta-analysis indicate that no treatment provided to men with pedohebephilic interest would qualify as an empirically-supported treatment. At present, while there is an evidence base for treatment change during interventions for pedohebephilic interest, the strongest conclusion about the body of research reviewed is there is evidence that treatment is effective, but the group of studies have limited internal validity. As outlined in Chapter 4, randomizing men to a treatment group and a no treatment or wait-list control group will allow for more confidence that intervention is responsible for any changes observed.

A second line of research will be to examine the relationship between change over the course of treatment with sexual recidivism. Such research may need to rely on archivally available phallometric data. However, such research is likely currently achievable given the availability of phallometric data. The findings of this research will extend the evidence base and potentially establish pedophilic interest as a psychologically meaningful risk factor (Mann et al., 2010).

6.4 Some Measures are Valid, While Others were not Supported

There is a large existing body of research that supports the predictive association between measures of pedophilic interest and sexual recidivism. The present research supports and extends this conclusion in finding that phallometric testing and the VRS-SO Sexual Deviance factor predicts sexual recidivism and, importantly, add incremental prediction to static risk. This latter finding is a relatively novel result, as the evidence base for incremental prediction on dynamic risk factors is less well established, particularly for phallometric tests of pedophilic interest. Incremental prediction provides strong evidence that phallometric testing and the VRS-SO

Sexual Deviance factor should continue to be included in clinical evaluations of sexual recidivism risk.

The results also provide validity evidence for latent structural models of pedophilic interest. A dimensional model received support as being predictive of sexual recidivism, while was not found to add incrementally to static risk ($p \approx .10$). This latter result suggests a potentially elevated risk for a Type II error and this analysis should be replicated, as there are few, if any, past attempts to examine incremental prediction using a continuous model. A two-level trichotomous model had relatively strong support, highlighting the importance of preferentiality of pedophilic interest when attempting to identify men at elevated risk of sexual recidivism. This trichotomous model likely performed better than a three-level model because there is little difference in sexual recidivism rates between non-pedophilic and non-preferentially pedophilic men.

Taken in conjunction with past research, there is strong support for preferential pedophilic interest as an empirically supported risk factor for sexual recidivism (Mann et al., 2010). *Empirically supported risk factors* have been found to predict sexual recidivism across three studies. The evidence indicates that preferential/exclusive pedophilic interest should be incorporated into clinical assessments of recidivism risk for men with histories of sexual offences against children and targeted in treatment programs aiming to reducing sexual recidivism. These results align well with the findings of the other research in this dissertation that preferentially pedophilic men show the most treatment change and that preferentially pedophilic men represent a class distinct from other men.

These results make intuitive sense, as men who are either preferentially or exclusively attracted to children have less motivation to establish intimate relationships with adults and

likely have a lower chance of maintaining such relationships if established. The absence of intimate relationships with adults is a distinct risk factor for sexual recidivism (Mann et al., 2010), one which may be expected to co-occur with preferential or exclusive pedophilic interest. A separate issue that may co-occur with preferential pedophilic interest is social isolation or loneliness. Men who experience preferential or exclusive pedophilic interest may feel isolated, not only due to a lack of an adult intimate partner, from others who they cannot reveal this aspect of their sexuality to and a sense of isolation from those towards whom they are attracted (i.e., children). Social isolation and loneliness are two additional risk factors that may co-occur with preferential or exclusive pedophilic interest. The taxometric findings also suggested that preferentially pedophilic men had elevated rates of sexual compulsivity. Paraphilic interest appears to have a relationship with hypersexuality/sexual compulsivity, and preferential pedophilic interest may also covary with these aspects of male sexuality. These speculations suggest that future research may profitably examine the psychosexual characteristics of preferentially pedophilic men to help further understand the issues that they experience.

A two-level trichotomous model may also have a practical advantage over a dimensional model. If a dimensional model continues to be validated as being robustly predictive of sexual recidivism, the consideration of how to use dimensionally operationalized pedophilic interest in clinical assessment invariably arises. The consideration regards identifying the point on dimensionally measured pedophilic interest men become higher risk to sexually re-offend. This is equivalent to creating cut scores along the pedophilic interest dimension. Latent dimensions do not easily lend themselves to identifying such cut scores (Ruscio et al., 2006). An advantage of a two-level trichotomous model is that an empirical cut-off is identifiable and can be used to identify an empirically and theoretically meaningful subset of men: those who are preferentially

attracted to children. This may lend a two-level trichotomous model more readily to clinical applications, such as assessment and treatment planning.

The results in Chapter 5 indicate pedophilic interest, when used to group men into crude pedophilic and non-pedophilic groups, does not reliably predict sexual recidivism. The present results are a replication of a result that has been found in several prior studies (Moulden et al., 2009; Wilson et al., 2011). At present, five studies (out of five located studies) indicate that when pedophilic interest is operationalized in a dichotomous manner, pedophilic interest is not associated with sexual recidivism. Taken together, the available research provides robust evidence that this method may not be a valid method of operationalizing pedophilic interest. This result may be due to whether men display pedophilic arousal is less informative than whether they have a high level of arousal to children and whether they have preferential arousal to children. A dichotomous operationalization potentially does not capture these nuances in male's sexual arousal.

Future research may extend these results in few directions. The first is to replicate the latent structural results in a larger and independent sample. Research may examine how best to combine tests indicative of pedophilic interest and phallometric tests indicative of preferential pedophilic interest, with established measures of static risk. Combining phallometric tests of pedophilic interest with static risk measures may provide more accurate estimates of sexual recidivism rates. While research shows that clinical adjustments to actuarial risk assessments tends to deteriorate predictive accuracy of such measures (Guay & Parent, 2018; Wormith, Hogg, & Guzzo, 2012), the present results provide some optimism that phallometric test scores can be used to inform risk assessment ratings in clinical evaluations of risk. Additional research is required to continue to identify how best to measure preferential/exclusive pedophilic interest.

Existing methods of operationalizing exclusivity in pedophilic interest do not provide a structured method of distinguishing between exclusively and non-exclusively pedophilic men (e.g., DSM-5 diagnostic criteria).

6.5 The Status of Phallometric Testing as a Measure of Pedophilic Interest

The three studies in this dissertation all assume the value and validity of phallometric tests for pedophilic interest. The status of phallometric testing as a measure of pedophilic interest and as a valid method to use in applied settings has been questioned for a long time and procedural innovations continue (Rosetti, Curry, Murphy, Bradford, & Fedoroff, 2019). Some commentators argue that men's sexual orientation is equal to their sexual arousal pattern (Bailey & Hsu, 2017), which may be used to support the importance of measuring men's arousal to individuals of different sexes and ages. The body of validity evidence for phallometric test scores being interpreted as indicative of a man's sexual orientation for sex and age (Rosenthal, Sylva, Safron, & Bailey, 2011; McPhail et al., 2017). The present results provided further evidence for the convergent and predictive validity of phallometric tests for pedophilic interest. A novel use of phallometric testing in the present research was using men's physiological arousal to children to identify latent structure in pedophilic interest. The second and third studies provide support for the taxometric results, as these studies support that grouping men according to their level of arousal to children, informed by taxometric results, leads to meaningful results in terms of treatment change and sexual recidivism. While the present results do not ameliorate the concerns regarding phallometric testing (Marshall & Fernandez, 2000), these findings do suggest phallometric tests may continue to be valuable in research.

6.6 Conclusion

Pedophilic interest is a central construct in understanding human sexuality and why some men commit sexual offences against children. The present dissertation advances knowledge on theoretical and applied issues with pedophilic interest in the following directions. Pedophilic interest may best be characterized as consisting of three-ordered classes of people: those who have no attraction to children, those who are non-preferentially attracted to children, and those who are preferentially attracted to children. These classes may be best understood as different in *kind* rather than different in *degree*. These empirical results have implications for how pedophilic interest is conceptualized, measured in research and clinical settings, and treated. For instance, preferentially pedophilic men were found to be at elevated risk for sexual reoffence, yet are able to manage their arousal to children after participating in behavioural or pharmacological interventions. Measures of pedophilic interest continue to accrue validity evidence for their ability to measure the construct and predict sexual recidivism. The results of this dissertation provide direction for future research into the nature pedophilic interest and its status as a construct central to our understanding of sexual offending against children.

References

Ahmed, A. O. (2010). *Differentiating classes from dimensions under unfavorable data conditions: Monte Carlo comparisons of taxometric and latent variable mixture models*, Unpublished doctoral dissertation, University of Southern Mississippi.

Alanko, K., Schmidt, A. F., Neutze, J., Bergen, E., Santtila, P., & Osterheider, M. (2017). Male sexual interest in and offending against children: The abused-abuser hypothesis. *Journal of Forensic Psychology Research and Practice*, *17*, 128–144. doi:10.1080/24732850.2017.1286544

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Association for the Treatment of Sexual Abusers. (2012). *Pharmacological interventions with adult male sexual offenders.* Beaverton, OR: Author.

Babchishin, K. M., Nunes, K. L., & Hermann, C. A. (2013). The validity of Implicit Association Test (IAT) measures of sexual attraction to children: A meta-analysis. *Archives of Sexual Behavior*, *42*, 487–499. doi:10.1007/s10508-012-0022-8

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*, 989–1000. doi:10.1037/abn0000213

Bailey, J. M., & Hsu, K. J. (2017). Orienting basic research on chronophilias. *Archives of Sexual Behavior*, *46*, 23–26. doi:10.1007/s10508-016-0885-1

Bailey, J. M., Hsu, K. J., & Bernhard, P. A. (2016). An Internet study of men sexually attracted to children: Sexual attraction patterns. *Journal of Abnormal Psychology*, *125*, 976–988. doi:10.1037/abn0000213

Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multimethod approach. *Criminal Justice and Behavior, 37,* 319–335. doi:10.1177/0093854809357598

Barsetti, I., Earls, C. M., Lalumière, M. L., & Bélanger, N. (1998). The differentiation of intrafamilial and extrafamilial heterosexual child molesters. *Journal of Interpersonal Violence, 13,* 275–286. doi:10.1177/088626098013002007

Baxter, D. J., Marshall, W. L., Barbaree, H. E., Davidson, P. R., & Malcolm, P. B. (1984). Deviant sexual behavior: Differentiating sex offenders by criminal and personal history, psychometric measures, and sexual responses. *Criminal Justice and Behavior, 11*, 477–501. doi:10.1177/0093854884011004007

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, *15*, 501–527. doi:10.1017/S0954579403000270

Becker, J. V., Kaplan, M. S., & Kavoussi, R. (1988). Measuring the effectiveness of treatment for the aggressive adolescent sexual offender. *Annals of the New York Academy of Sciences, 528,* 215–222. doi:10.1111/j.1749-6632.1988.tb50865.x

Becker, J. V., Stein, R. M., Kaplan, M. S., & Cunningham-Rathner, J. (1992). Erection response characteristics of adolescent sex offenders. *Annals of Sex Research, 5*, 81–86. doi:10.1177/107906329200500202

Beggs, S. M., & Grace, R. C. (2010). Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse: Journal of Research and Treatment, 22,* 234 –251. doi:10.1177/1079063210369014

Beier, K. M. (1998). Differential typology and prognosis for dissexual behavior–a follow-up study of previously expert-appraised child molesters. *International Journal of Legal Medicine, 111,* 133–141. doi:10.1007/s004140050133

Biostat. (2007). Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Author.

Blanchard, R. (2018). Fraternal birth order, family size, and male homosexuality: Meta-analysis of studies spanning 25 years. *Archives of Sexual Behavior, 47,* 1–15. doi:10.1007/s10508-017-1007-4

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,* 118–126. doi:10.1037/1040-3590.13.1.118

Blanchard, R., Kuban, M. E., Blak, T., Cantor, J. M., Klassen, P. E., & Dickey, R. (2009) Absolute versus relative ascertainment of pedophilia in men. *Sexual Abuse: A Journal of Research and Treatment, 21,* 431–441. doi:10.1177/1079063209347906

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,* 13–29. doi:10.1007/s10508-010-9675-3

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.) New York: Routledge.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, England: Wiley.

Borsboom, D., Rhemtulla, M., Cramer, A. O. J., van der Maas, H. L. J., Scheffer, M., & Dolan, C. V. (2016). Kinds versus continua: A review of psychometric approaches to uncover the

structure of psychiatric constructs. *Psychological Medicine, 46,* 1567–1579. doi:10.1017/S0033291715001944

Bouchard, K. N., Dawson, S. J., & Lalumière, M. L. (2017). The effects of sex drive and paraphilic interests on paraphilic behaviours in a nonclinical sample of men and women. *Canadian Journal of Human Sexuality, 26,* 97–111. doi:10.3138/cjhs.262-a8

Bradford, J. M., & Pawlak, A. (1993). Effects of cyproterone acetate on sexual arousal patterns of pedophiles. *Archives of Sexual Behavior, 22,* 629–641. doi:10.1007/BF01543305

Brankley, A. E. (2019). *A taxometric analysis of pedophilia in adult males convicted of sexual offences: Evidence for a taxon* [Unpublished doctoral dissertation]. Ryerson University.

Briken, P., & Kafka, M. P. (2007). Pharmacological treatments for paraphilic patients and sexual offenders. *Current Opinion in Psychiatry, 20,* 609–613. doi:10.1097/YCO.0b013e3282f0eb0b

Burda, B. U., O' Connor, E. A., Webber, E. M., Redmond, N., & Perdue, L. A. (2017). Estimating data from figures with a Web-based program: Considerations for a systematic review. *Research Synthesis Methods, 8,* 258–262. doi:10.1002/jrsm.1232

Canales, D. D., Olver, M. E., & Wong, S. C. (2009). Construct validity of the Violence Risk Scale—Sexual Offender Version for measuring sexual deviance. *Sexual Abuse: A Journal of Research and Treatment, 21,* 474–492. doi:10.1177/1079063209344990

Cantor, J. M. (2018). Can pedophiles change? *Current Sexual Health Reports, 10,* 203–206. doi:10.1007/s11930-018-0165-2

Cantor, J. M., & Blanchard, R. (2012). White matter volumes in pedophiles, hebephiles, and teleiophiles. *Archives of Sexual Behavior, 41,* 749–752. doi:19.1007/s10508-012-9954-2

Cantor, J. M., Kabami, N., Christensen, B. K., Zipursky, R. B., Barbaree, H. E., Dickey, R., . . . Blanchard, R. (2008). Cerebral white matter deficiencies in pedophilic men. *Journal of Psychiatric Research, 42*, 167–183. doi:10.1016/j.psychires.2007.10.013

Cantor, J. M., Klein, C., Lykins, A., Rullo, J. E., Thaler, L., & Walling, B. R. (2013). A treatment-oriented typology of self-identified hypersexuality referrals. *Archives of Sexual Behavior, 42*, 883–893. doi:10.1007/s10508-013-0085-1

Cantor, J. M., Lafaille, S., Soh, D. W., Moayendi, M., Mikulus, D. J., & Girard, T.A. (2015). Diffusion tensor imaging of pedophilia. *Archives of Sexual Behavior, 44*, 2161–2172. doi:1.1007/s10508-015-0629-7

Cantor, J. M., & McPhail, I. V. (2015). Sensitivity and specificity of the phallometric test for hebephilia. *Journal of Sexual Medicine, 12,* 1940–1950. doi:10.1111/jsm.12970

Carvalho, J., Stulhofer, A., Vieira, A. L., & Jurin, T. (2015). Hypersexuality and high sexual desire: Exploring the structure of problematic sexuality. *Journal of Sexual Medicine, 12,* 1356–1367. doi:10.1111/jsm.12865

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Erlbaum.

Cohen, L. J., & Galynker, I. I. (2002). Clinical features of pedophilia and implications for treatment. *Journal of Psychiatric Practice, 8*, 276–289. doi:10.1097/00131746-200209000-00004

Cotter, A., & Beaupré, P. (2014). Police reported sexual offences against children and youth in Canada, 2012. *Juristat*, Statics Canada Catalogue no. 85–002-X. Ottawa, Ontario, Canada: Statistics Canada.

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences, 26,* 364–368. doi:10.1017/S2045796016000809

Davis, K. (2017). *The relation of childhood abuse experiences to problematic sexual behaviors in male youths who have sexually offended,* Unpublished doctoral dissertation, Brandeis University.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Dyer, T. J., & Olver, M. E. (2016). Self-reported psychopathy and its association with deviant sexual fantasy and sexual compulsivity in a nonclinical sample. *Sexual Offender Treatment, 11*, 1–18.

Dyshniku, F., Murray, M. E., Fazio, R. L., Lykins, A. D., & Cantor, J. M. (2015). Minor physical anomalies as a window into the prenatal origins of pedophilia. *Archives of Sexual Behavior, 44*, 2151–2159. doi:10.1007/s10508-015-0564-7

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634. doi:10.1136/bmj.315.7109.629

Eher, R., Olver, M. E., Heurix, I., Schilling, F., & Rettenberger, M. (2015). Predicting reoffense in pedophilic child molesters by clinical diagnoses and risk assessment. *Law and Human Behavior, 39*, 571–580. doi:10.1037/lhb0000144

Eher, R., Rettenberger, M., Matthes, A., & Schilling, F. (2010). Stable dynamic risk factors in child sexual abusers: The incremental predictive power of narcissistic personality traits

beyond the Static-99/Stable-2007 priority categories on sexual reoffense. *Sexual Offender Treatment, 5,* 1–12.

Fazio, R. L., Dyshniku, F., Lykins, A. D., & Cantor, J. M. (2017). Leg length versus torso length in pedophilia further evidence of atypical physical development early in life. *Sexual Abuse: A Journal of Research and Treatment, 29,* 500–514. doi:1079063215609936

Fazio, R. L., Lykins, A. D., & Cantor, J. M. (2014). Elevated rates of atypical handedness in paedophilia: Theory and implications. *Laterality: Asymmetries of Body, Brain and Cognition, 19,* 690–704. doi:10.1080/1357650X.2014.898648

Fedoroff, J. P. (2018). Can People with Pedophilia Change?: Yes they can! *Current Sexual Health Reports, 10,* 207–212. doi:10.1007/s11930-018-0166-1

Finkelhor, D., & Araji, S. (1986). Explanations of pedophilia: A four factor model. *Journal of Sex Research, 22,* 145–161. doi:10.1080/00224498609551297

Furukawa, T. A., Weitz, E. S., Tanaka, S., Hollon, S. D., Hofmann, S. G., Andersson, G., ... & Mergl, R. (2017). Initial severity of depression and efficacy of cognitive–behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. *The British Journal of Psychiatry, 210,* 190–196. doi:10.1192/bjp.bp.116.187773

Gannon, T. A., Olver, M. E., Mallion, J. S., & James, M. (2018). *Does specialized psychological treatment for offending reduce recidivism? A meta-analysis examining staff and program variables as predictors of treatment effectiveness.* Manuscript submitted for publication.

Garcia, F. D., & Thibaut, F. (2011). Current concepts in the pharmacotherapy of paraphilias. *Drugs, 71,* 771–790. doi:10.2165/11585490-000000000-00000

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480. doi:10.1037/0022-3514.74.6.1464

Guay, J. P., & Parent, G. (2018). Broken legs, clinical overrides, and recidivism risk: An analysis of decisions to adjust risk levels with the LS/CMI. *Criminal Justice and Behavior, 45,* 82–100. doi:10.1177/0093854817719482

Hall, G. C. N., & Hirschman, R. (1992). Sexual aggression against children: A conceptual perspective of etiology. *Criminal Justice and Behavior, 19,* 8–23. doi:10.1177/0093854892019001003

Hanson, R. K. (2010). Dimensional measurement of sexual deviance. *Archives of Sexual Behavior, 39,* 401–404. doi:10.1007/s10508-009-9575-6

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*. Ottawa, ON: Public Safety and Emergency Preparedness Canada.

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,* 1154–1163. doi:10.1037/0022-006X.73.6.1154

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1–21. doi:10.1037/a0014421

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. doi:10.1023/A:1005482921 333.

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine, 42,* 903–920. doi:10.1017/S0033291711001966

Helmus, L., Ó Ciardha, C., & Seto, M. C. (2014). The Screening Scale for Pedophilic Interests (SSPI): Construct, predictive, and incremental validity. *Law and Human Behavior, 39,* 35–43. doi:10.1037/lhb0000099

Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior, 39,* 1148–1171. doi:10.1177/0093854812443648

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: Journal of Research and Treatment, 24,* 64–101. doi:10.1177/1079063211409951

Higgins, J., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. doi:10.1136/bmj.327.7414.557

Hoffmann, H. (2017). Situating human sexual conditioning. *Archives of Sexual Behavior, 46,* 2213–2229. doi:10.1007/s10508-017-1030-5

Howes, R. J. (1995). A survey of plethysmographic assessment in North America. *Sexual Abuse: A Journal of Research and Treatment, 7,* 9–24. doi:10.1007/BF02254871

Hunter, J. A., & Goodwin, D. W. (1992). The clinical utility of satiation therapy with juvenile sexual offenders: Variations and efficacy. *Annals of Sex Research, 5,* 71–80. doi:10.1007/BF00849732

Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal, 176,* 1091–1096. doi:10.1503/cmaj.060410

Janssen, E. (Ed.). (2007). *The psychophysiology of sex*. Bloomington, IN: Indiana University Press.

Kafka, M. P. (2010). Hypersexual disorder: A proposed diagnosis for DSM-V. *Archives of Sexual Behavior, 39,* 377–400. doi:10.1007/s10508-009-9574-7

Kafka, M. P., & Hennen, J. (2003). Hypersexual desire in males: Are males with paraphilias different from males with paraphilia-related disorders? *Sexual Abuse: A Journal of Research and Treatment, 15,* 307–321. doi:10.1023/A:1025000227956

Kamenov, K., Twomey, C., Cabello, M., Prina, A. M., & Ayuso-Mateos, J. L. (2017). The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: a meta-analysis. *Psychological Medicine, 47,* 414–425. doi:10.1017/S0033291716002774

Kelly, R. J. (1982). Behavioral reorientation of pedophiliacs: Can it be done?. *Clinical Psychology Review, 2,* 387–408. doi:10.1016/0272-7358(82)90020-4

Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparison for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 285–299. doi:10.1037//0022-006X.67.3.285

King, M. (2010). *Neurodevelopmental vulnerability for pedophilia: A taxometric investigation of its latent structure* [Unpublished doctoral dissertation]. University of Maine.

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,* 423–436. doi:10.1007/s10508-006-9091-x

Klein, V., Schmidt, A. F., Turner, D., & Briken, P. (2015). Are sex drive and hypersexuality associated with pedophilic interest and child sexual abuse in a male community sample? *PloS One*, *10*, e0129730. doi:10.1371/journal.pone.0129730

Knight, R. A., & Graham, F. J. (2017). Hypersexuality: Equifinal, cohesive, clinical presentation or symptom cluster with multiple underlying mechanisms? *Archives of Sexual Behavior, 46,* 2261–2264. doi:10.1007/s10508-017-1089-z

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8,* 355–362. doi:10.1177/1948550617697177

Långström, N., & Hanson, R. K. (2006). High rates of sexual behavior in the general population: Correlates and predictors. *Archives of Sexual Behavior*, *35*, 37–52. doi:10.1007/s10508-006-8993-y

Långström, N., & Seto, M. C. (2006). Exhibitionistic and voyeuristic behavior in a Swedish national population survey. *Archives of Sexual Behavior*, *35*, 427–435. doi:10.1007/s10508-006-9042-6

Långström, N., & Zucker, K. J. (2005). Transvestic fetishism in the general population. *Journal of Sex & Marital Therapy*, *31*, 87–95. doi:10.1080/00926230590477934

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). London: Wiley-Blackwell.

Looman, J., & Marshall, W. L. (2001). Phallometric assessments designed to detect arousal to children: The responses of rapists and child molesters. *Sexual Abuse: A Journal of Research and Treatment, 13,* 3–13. doi:10.1177/107906320101300102

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., ... & Krause, E. D. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice, 9,* 2–12. doi:10.1093/clipsy.9.1.2

Mackaronis, J. E., Byrne, P., Strassberg, D., Marcus, D., & Solari, J. (2011, November). *What does phallometry assess: Categories or continua?* Presentation given at the Conference of the Association for the Treatment of Sexual Abusers, Toronto.

Mackaronis, J. E., Strassberg, D. S., & Marcus, D. K. (2011). The latent structure of Multiphasic Sex Inventory–assessed pedophilic interest. *Psychological Assessment, 23*, 1017–1022. doi:10.1037/a0024625

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: A Journal of Research and Treatment, 22,* 191–217. doi:10.1177/1079063210366039

Marshall, W.L., Anderson, D., & Fernandez, Y.M. (1999). *Cognitive behavioural treatment of sexual offenders.* Chichester, UK: John Wiley & Sons.

Marshall, W. L., & Fernandez, Y. M. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychological Review*, *20*, 807-822. doi:10.1016/S0272-7358(99)00013-6

Marshall, W. L., O' Brien, M. D. & Marshall, L. E. (2009). Modifying sexual preferences. In A.R. Beech, L. A. Craig, & K. D. Browne (Eds.), *Assessment and treatment of sex offenders. A handbook* (pp. 311–327). Chichester, UK: Wiley.

McAleavey, A. A., Youn, S. J., Xiao, H., Castonguay, L. G., Hayes, J. A., & Locke, B. D. (2017). Effectiveness of routine psychotherapy: Method matters. *Psychotherapy Research, 29,* 139–156. doi:10.1080/10503307.2017.1395921

McEvoy, P. M. & Nathan, P. (2007). Effectiveness of cognitive behavior therapy for diagnostically heterogeneous groups: A benchmarking study. *Journal of Consulting and Clinical Psychology, 75,* 344–350. doi:10.1037/0022-006X.75.2.344

McGrath, R. E. (2008). Inferential errors in taxometric analyses of ordered three-class constructs. *Journal of Personality Assessment, 90,* 11–25. doi:10.1080/00223890701356755

McGrath, R. J., Cumming, G. F., Burchard, B. L., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management.* Brandon, VT: Safer Society Press.

McPhail, I. V. (2018). Age of onset in pedohebephilic interests [Letter to the Editor]. *Archives of Sexual Behavior, 47,* 1313–1317. doi:10.1007/s10508-018-1198-3

McPhail, I. V., & Cantor, J. M. (2015). Pedophilia, height, and the magnitude of the association: A research note. *Deviant Behavior, 36,* 288–292. doi:10.1080/01639625.2014.935644

McPhail, I. V., Hermann, C. A., Fernane, S., Fernandez, Y. M., Nunes, K. L., & Cantor, J. M. (2017). Validity in phallometric testing for sexual interests in children: A meta-analytic review. *Assessment*. Advance online publication. doi:1073191117706139

McPhail, I. V., & Olver, M. E. (2018). *Effectiveness of treatments for pedohebephilic interest in sexual offenders against children: A meta-analytic review.* Manuscript in preparation.

McPhail, I. V., Olver, M. E., Brouillette-Alarie, S., & Looman, J. (2018). Taxometric analysis of the latent structure of pedophilic interest. *Archives of Sexual Behavior, 47,* 2223–2240. doi:10.1007/s10508-018-1225-4

McPhail, I. V., & Stephens, S. (2019). *Unpublished data from an online sample of pedohebephilic and teleiophilic men.*

Meehl, P. E. (1977). Specific etiology and other forms of strong influence: Some quantitative meanings. *The Journal of Medicine and Philosophy, 2,* 33–53. doi:10.1093/jmp/2.1.33

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality, 60,* 117–174. doi:10.1111/j.1467-6494.1992.tb00269.x

Meehl, P. E. (1995a). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist, 50,* 266–275. doi:10.1037/0003-066X.50.4.266

Meehl, P. E. (1995b). Extension of the MAXCOV–HITMAX taxometric procedure to situations of sizeable nuisance covariance. In D. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 81–92). Palo Alto, CA: Davies-Black.

Meehl, P. E. (2004). What's in a taxon? *Journal of Abnormal Psychology, 113,* 39–43. doi:10.1037/0021-843X.113.1.39

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74,* 1059–1274.

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator

(MAXCOV procedure). *Psychological Reports, 78*, 1091–1227. doi:10.2466/pr0.1996.78.3c.1091

Milner, R. J. (2016). *Desistance in men who have previously committed sexual offences: An exploration of the early processes*. Unpublished doctoral dissertation, University of York.

Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. J. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality and Quantity, 42*, 513–525. doi:10.1007/s11135-006-9057-z

Moulden, H. M., Firestone, P., Kingston, D., & Bradford, J. (2009). Recidivism in pedophiles: An investigation using different diagnostic methods. *Journal of Forensic Psychiatry & Psychology, 20,* 680–701. doi:10.1080/14789940903174055

Müller, K., Curry, S., Ranger, R., Briken, P., Bradford, J., & Fedoroff, J. P. (2014). Changes in sexual arousal as measured by penile plethysmography in men with pedophilic sexual interest. *Journal of Sexual Medicine, 11*, 1221–1229. doi:10.1111/jsm.12488

Nota, N. M., Wiepjes, C. M., de Blok, C. J., Gooren, L. J., Kreukels, B. P., & den Heijer, M. (2019). The occurrence of acute cardiovascular events in transgender individuals receiving hormone therapy: Results from a large cohort study. *Circulation*. Advance online publication. doi: 10.1161/CIRCULATIONAHA.118.038584

Ó Ciardha, C., Attard-Johnson, J., & Bindemann, M. (2018). Latency-based and psychophysiological measures of sexual interest show convergent and concurrent validity. *Archives of Sexual Behavior, 47,* 637–649. doi:10.1007/s10508-017-1133-z

Olver, M. E., & Eher, R. (2019). *Predictive properties and factor structure of the VRS-SO: Results from an Austrian sample.* Manuscript under review.

Olver, M. E., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C. (2014). A psychometric examination of treatment change in a multisite sample of treated Canadian federal sexual offenders. *Law and Human Behavior, 38,* 544–559. doi:10.1037/lhb0000086

Olver, M. E., Neumann, C. S., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C.P. (2018). Construct validity of the Violence Risk Scale-Sexual Offender version instrument in a multisite sample of treated sexual offenders. *Assessment, 25,* 40-55. doi:10.1177/1073191116643819

Olver, M. E., & Wong, S. C. (2006). Psychopathy, sexual deviance, and recidivism among sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 18,* 65–82. doi:10.1177/107906320601800105

Olver, M. E., Wong, S. C., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19,* 318–329. doi:10.1037/1040-3590.19.3.318

Owen, M. J., O'donovan, M. C., Thapar, A., & Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. *The British Journal of Psychiatry, 198,* 173–175. doi:10.1192/bjp.bp.110.084384

Paolucci, E. O., Genuis, M. L., & Violato, C. (2001). A meta-analysis of the published research on the effects of child sexual abuse. *The Journal of Psychology, 135,* 17–36. doi:10.1080/00223980109603677

Prentky, R. A. (1997). Arousal reduction in sexual offenders: A review of antiandrogen interventions. *Sexual Abuse: A Journal of Research and Treatment, 9,* 335–347. doi:10.1007/BF02674857

Quinsey, V. L., & Chaplin, T. C. (1988). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence, 3,* 259–274. doi:10.1177/088626088003003001

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d,* and *r*. *Law and Human Behavior, 29,* 615–620. doi:10.1007/s10979-005-6832-7

Ricci, R. J., Clayton, C. A., & Shapiro, F. (2006). Some effects of EMDR on previously abused child molesters: Theoretical reviews and preliminary findings. *The Journal of Forensic Psychiatry & Psychology*, *17*, 538–562. doi:10.1080/14789940601070431

Rohatgi, A. (2018). *WebPlotDigitizer* (Version 3.8) [Computer software]. Austin, Texas: Author. Available from https://automeris.io/WebPlotDigitizer [accessed September 1, 2017].

Rommelse, N. N., Franke, B., Geurts, H. M., Hartman, C. A., & Buitelaar, J. K. (2010). Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *European Child & Adolescent Psychiatry*, *19*, 281–295. doi:10.1007/s00787-010-0092-x

Rosenthal, A. M., Sylva, D., Safron, A., & Bailey, J. M. (2012). The male bisexuality debate revisited: Some bisexual men have bisexual arousal patterns. *Archives of Sexual Behavior*, *41*, 135–147. doi:10.1007/s10508-011-9881-7

Rosenthal, R. (1993). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage Publications.

Rosetti, L., Curry, S., Murphy, L., Bradford, J. B., & Fedoroff, J. P. (2019). A comparison of sexual arousal in men exposed to visual stimuli with and without facial blurring. *Sexual Abuse*. Advance online publication. doi:1079063219828784.

Ruscio, J. (2014). *Taxometric programs for the R computing environment: User's manual.* Retrieved from ruscio.pages.tcnj.edu.

Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). *Introduction to the taxometric method: A practical guide.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ruscio, J., & Kaczetow, W. (2009). Differentiating categories and dimensions: Evaluating the robustness of taxometric analysis. *Multivariate Behavioral Research, 44,* 259–280. doi:10.1080/00273170902794248

Ruscio, J., & Ruscio, A. M. (2004). A nontechnical introduction to the taxometric method. *Understanding Statistics, 3,* 151–194. doi:10.1207/s15328031us0303_2

Ruscio, J., Ruscio, A. M., & Meron, M. (2007). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research, 42,* 349–386. doi:10.1080/00273170701360795

Ruscio, J., Walters, G. D., Marcus, D. K., & Kaczetow, W. (2010). Comparing the relative fit of categorical and dimensional latent variable models using consistency tests. *Psychological Assessment, 22,* 5–21. doi:10.1037/a0018259

Santtila, P., Mokros, A., Hartwig, M., Varjonen, M., Jern, P., Witting, K., ... & Sandnabba, N. K. (2010). Childhood sexual interactions with other children are associated with lower preferred age of sexual partners including sexual interest in children in adulthood. *Psychiatry Research, 175,* 154–159. doi:10.1016/j.psychres.2008.10.021

Schmidt, A. F., Babchishin, K. M., & Lehmann, R. J. (2017). A meta-analysis of viewing time measures of sexual interest in children. *Archives of Sexual Behavior, 46,* 287–300. doi:10.1007/s10508-016-0806-3

Schmidt, A. F., Gykiere, K., Vanhoeck, K., Mann, R. E., & Banse, R. (2014). Direct and indirect measures of sexual maturity preferences differentiate subtypes of child sexual abusers. *Sexual Abuse: A Journal of Research and Treatment, 26,* 107–128. doi:10.1177/1079063213480817

Schmidt, A. F., Mokros, A., & Banse, R. (2013). Is pedophilic sexual preference continuous? A taxometric analysis based on direct and indirect measures. *Psychological Assessment, 25,* 1146–1153. doi:10.1037/a0033326

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. Zeitschrift für Psychologie/Journal of Psychology, 215, 90–103. doi:10.1027/0044-3409.215.2.90.

Seto, M. C. (2012). Is pedophilia a sexual orientation? *Archives of Sexual Behavior, 41,* 231–236. doi:10.1007/s10508-011-9882-6

Seto, M. C. (2017). The puzzle of male chronophilias. *Archives of Sexual Behavior, 46,* 3–22. doi:10.1007/s10508-016-0799-y

Seto, M. C. (2018). *Pedophilia and sexual offending against children: Theory, assessment, and intervention* (2nd ed.). Washington, DC: American Psychological Association.

Seto, M. C., Cantor, J. M., & Blanchard, R. (2006). Child pornography offenses are a valid diagnostic indicator of pedophilia. *Journal of Abnormal Psychology, 115,* 610–615. doi: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,* 455–466. doi: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: A Journal of Research and Treatment, 13,* 15–25. doi:10.1177/107906320101300103

Seto, M. C., Lalumière, M. L., & Kuban, M. (1999). The sexual preferences of incest offenders. *Journal of Abnormal Psychology, 108,* 267–272. doi: 10.1037//0021-843X.108.2.267

Seto, M. C., Sandler, J. C., & Freeman, N. J. (2017). The revised Screening Scale for Pedophilic Interests (SSPI-2): Concurrent and predictive validity. *Sexual Abuse: A Journal of Research and Treatment, 29,* 636-657. doi:10.1177/1079063215618375

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 47,* 921–929. doi:10.1097/CHI.0b013e318179964f

Skorska, M. N., & Bogaert, A. F. (2016). Sexual orientation, objective height, and self-reported height. *The Journal of Sex Research, 54,* 19–32. doi:10.1080/00224499.2015.1124831

Skorska, M. N., & Bogaert, A. F. (2017). Pubertal stress and nutrition and their association with sexual orientation and height in the Add Health data. *Archives of Sexual Behavior*, *46,* 217–236. doi:10.1007/s10508-016-0800-9

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,* 585–595. doi:10.1037/ccp0000194

Stephens, S., Leroux, E., Skilling, T., Cantor, J. M., & Seto, M. C. (2017). Taxometric analyses of pedophilia utilizing self-report, behavioral, and sexual arousal indicators. *Journal of Abnormal Psychology, 126,* 1114–1119. doi:10.1037/abn0000291

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Carpenter, J. R. (2016a). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ, 355,* i4919. doi:10.1136/bmj.i4919

Sterne, J. A. C., Higgins, J. P. T., Elbers, R. G., Reeves, B. C., & the development group for ROBINS-I. (2016b). Risk of Bias in non-randomized studies of interventions (ROBINS-I): Detailed guidance, updated 12 October 2016. Available from http://www.riskofbias.info [accessed February 20, 2017].

Stulhofer, A., Jurin, T., & Briken, P. (2016). Is high sexual desire a facet of male hypersexuality? Results from an online study. *Journal of Sex and Marital Therapy, 42,* 665–680. doi:10.1080/0092623X.2015.1113585

Sutton, K. S., Stratton, N., Pytyck, J., Kolla, N. J., & Cantor, J. M. (2015). Patient characteristics by type of hypersexuality referral: A quantitative chart review of 115 consecutive male cases. *Journal of Sex & Marital Therapy*, *41*, 563–580. doi:10.1080/0092623X.2014.93553

Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Tozdan, S., & Briken, P. (2015). The earlier, the worse? Age of onset of sexual interest in children. *The Journal of Sexual Medicine*, *12,* 1602–1608. doi:10.1111/jsm.12927

Turner, D., & Briken, P. (2018). Treatment of paraphilic disorders in sexual offenders or men with a risk of sexual offending with luteinizing hormone-releasing hormone agonists: An updated systematic review. *The Journal of Sexual Medicine, 15,* 77–93. doi:10.1016/j.jsxm.2017.11.013.

van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment, 30,* 179–191. doi:10.1037/pas0000454

van Wieringen, K., & Cribbie, R. A. (2014). Evaluating clinical significance: Incorporating robust statistics with normative comparison tests. *British Journal of Mathematical and Statistical Psychology, 67,* 213–230. doi:10.1111/bmsp.12015

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua.* Thousand Oaks, CA: Sage.

Walters, G. D., McGrath, R. E., & Knight, R. A. (2010). Taxometrics, polytomous constructs, and the comparison curve fit index: A Monte Carlo analysis. *Psychological Assessment, 22,* 149–156. doi:10.1037/a0017819

Walton, M. T., Cantor, J. M., Bhullar, N., & Lykins, A. D. (2017). Hypersexuality: A critical review and introduction to the "sexhavior cycle". *Archives of Sexual Behavior, 46,* 2231–2251. doi:10.1007/s10508-017-0991-8juroin

Wang, C. T., & Holton, J. (2007). *Total estimated cost of child abuse and neglect in the United States*. Chicago, IL: Prevent Child Abuse America.

Ward, T., & Beech, A. R. (2004). The etiology of risk: A preliminary model. *Sexual Abuse: A Journal of Research and Treatment, 16,* 271–284. doi:10.1023/B:SEBU.0000043323.75071.5f

Ward, T., & Beech, A. R. (2006). An integrated theory of sexual offending. *Aggression and Violent Behavior, 11,* 44–63. doi:10.1016/j.avb.2005.05.002

Ward, T., & Siegert, R. J. (2002). Toward and comprehensive theory of child sexual abuse: A theory knitting perspective. *Psychology, Crime, and Law, 9,* 319–351. doi:10.1080/10683160208401823

Whitaker, W., Le, B., Hanson, R. K., Baker, C. K., McMahon, P. M., Ryan, G., . . . Rice, D. D. (2008). Risk factors for the perpetration of child sexual abuse: A review and meta-analysis. *Child Abuse & Neglect, 32,* 529–548. doi:10.1016/j.chiabu.2007.08.005

Wilcox, R. R. (2005). New methods for comparing groups: strategies for increasing the probability of detecting true differences. *Current Directions in Psychological Science, 14*, 272–275. doi:10.1111/j.0963-7214.2005.00379.x

Wilson, R. J., Abracen, J., Looman, J., Picheca, J. E., & Ferguson, M. (2011). Pedophilia: An evaluation of diagnostic and risk prediction methods. *Sexual Abuse: Journal of Research and Treatment, 23,* 260–274. doi:10.1177/1079063210384277

Wong, S. C. P., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. E. (2003–2017). *The Violence Risk Scale-Sexual Offender Version (VRS-SO)*. Saskatoon, Saskatchewan, Canada: Regional Psychiatric Centre and University of Saskatchewan.

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior, 39,* 1511–1538. doi:10.1177/0093854812455741

Zauche, L. H., Mahoney, A. E. D., & Higgins, M. K. (2017). Predictors of co-occurring neurodevelopmental disabilities in children with Autism Spectrum Disorders. *Journal of Pediatric Nursing, 35*, 113–119. doi:10.1016/j.pedn.2017.04.002

Appendix A: Supplemental Figures from Taxometric Analyses

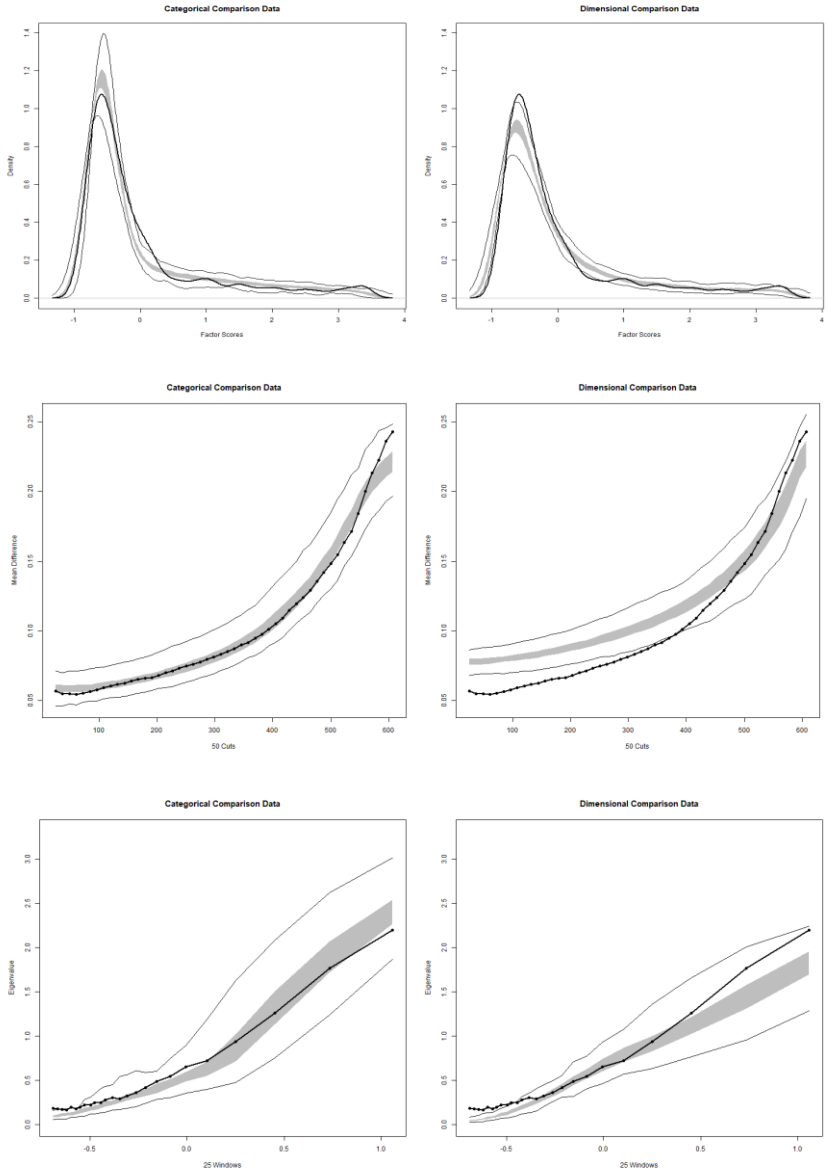

Figure A-1. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes in the Philippe-Pinel dataset.

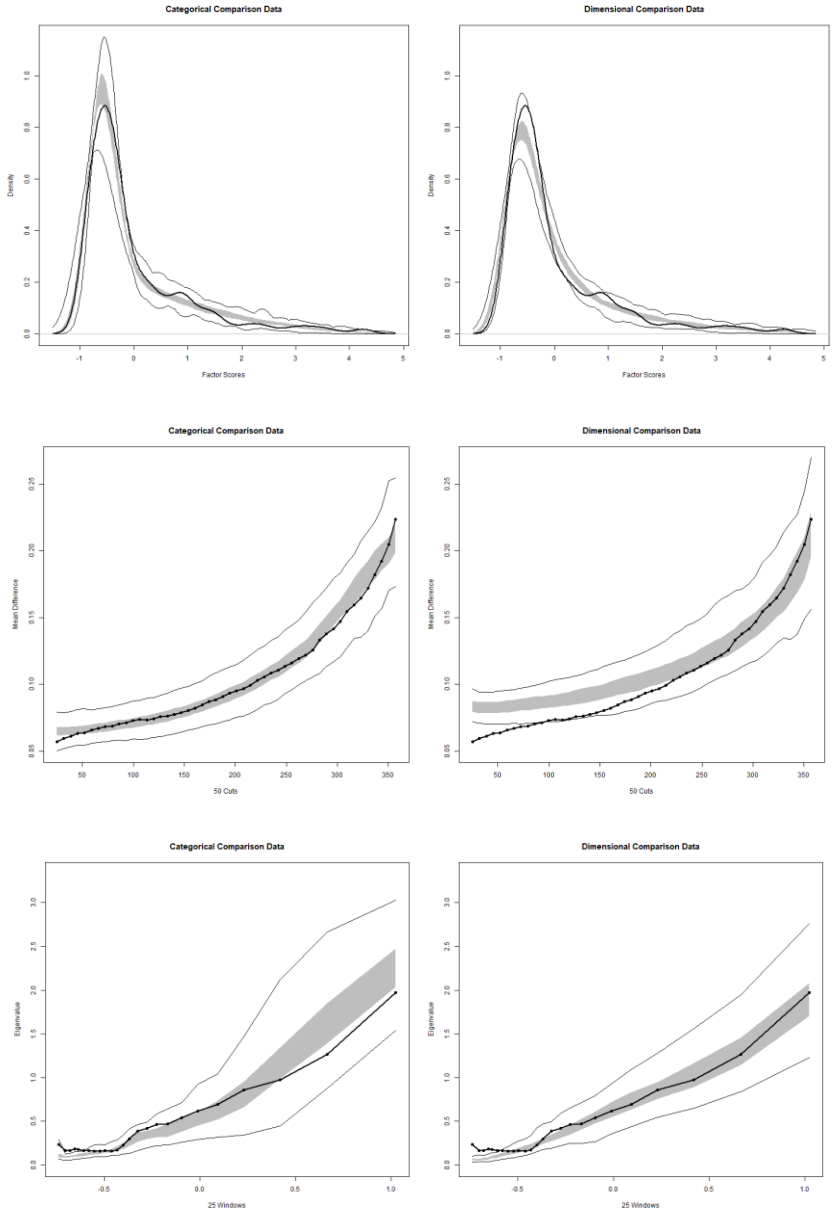

Figure A-2. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the RTC 1: Audio dataset.

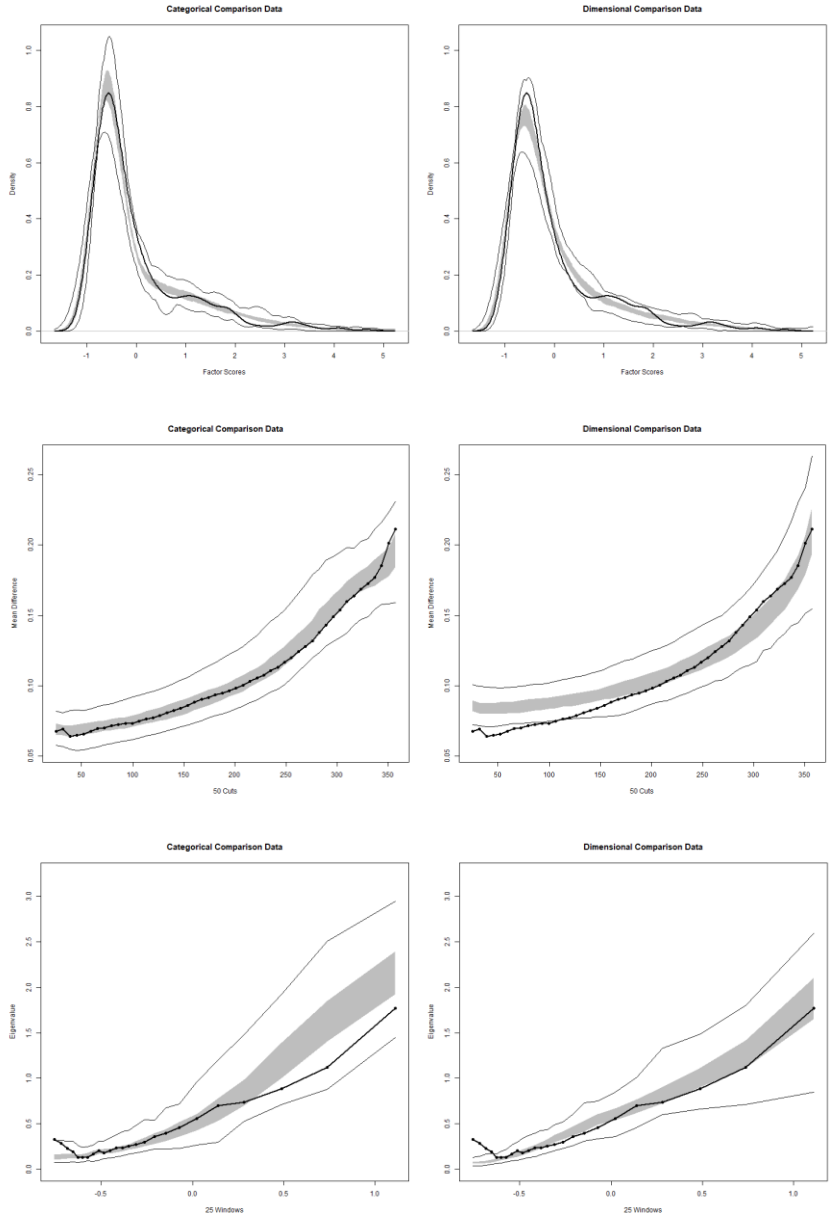

Figure A-3. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the RTC 1: Audio dataset.

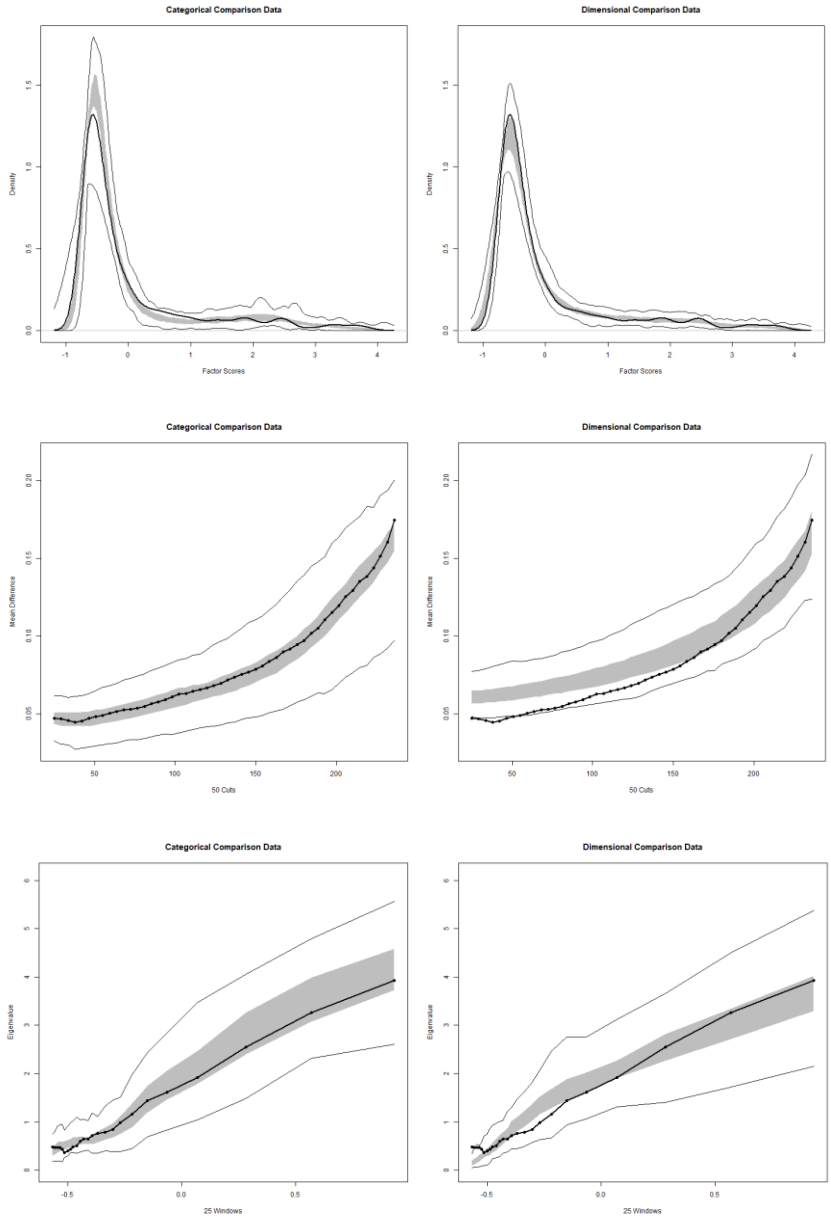

Figure A-4. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes in the RPC dataset.

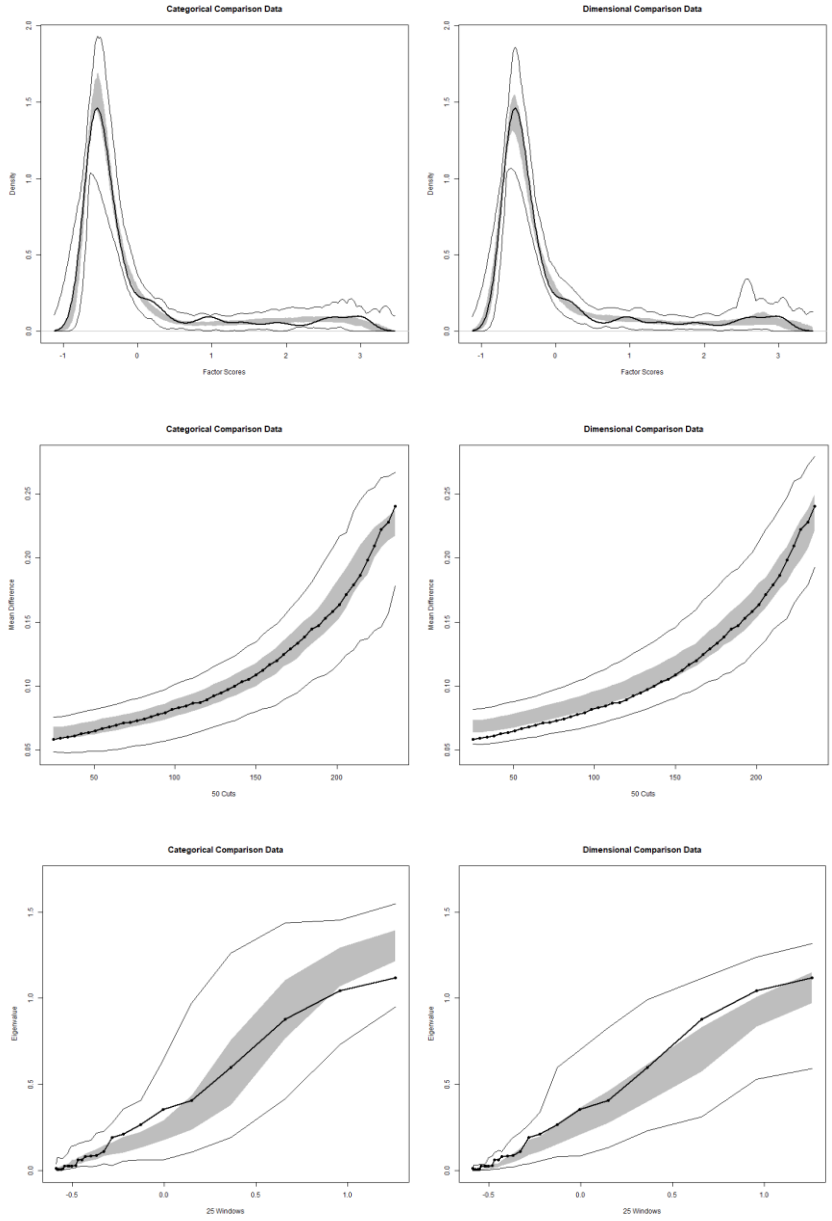

Figure A-5. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the RPC dataset.

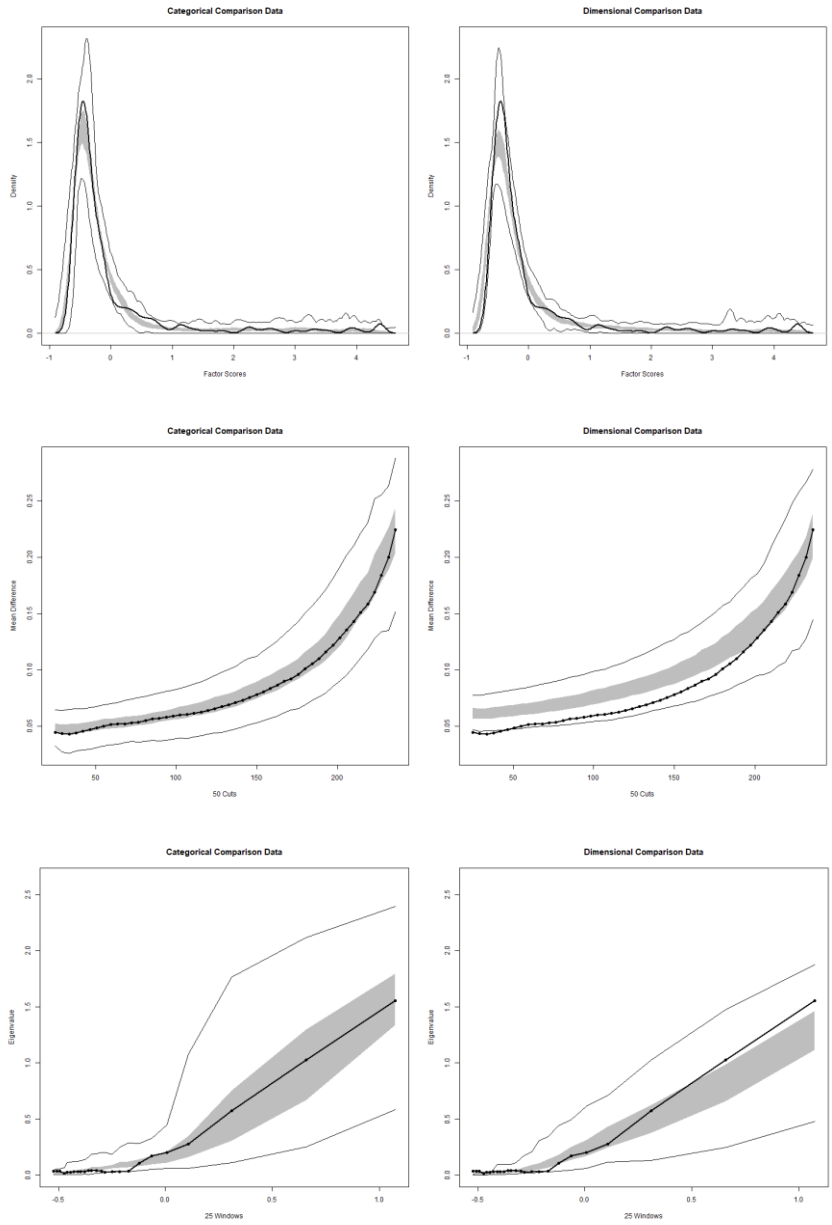

Figure A-6. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the RPC dataset.

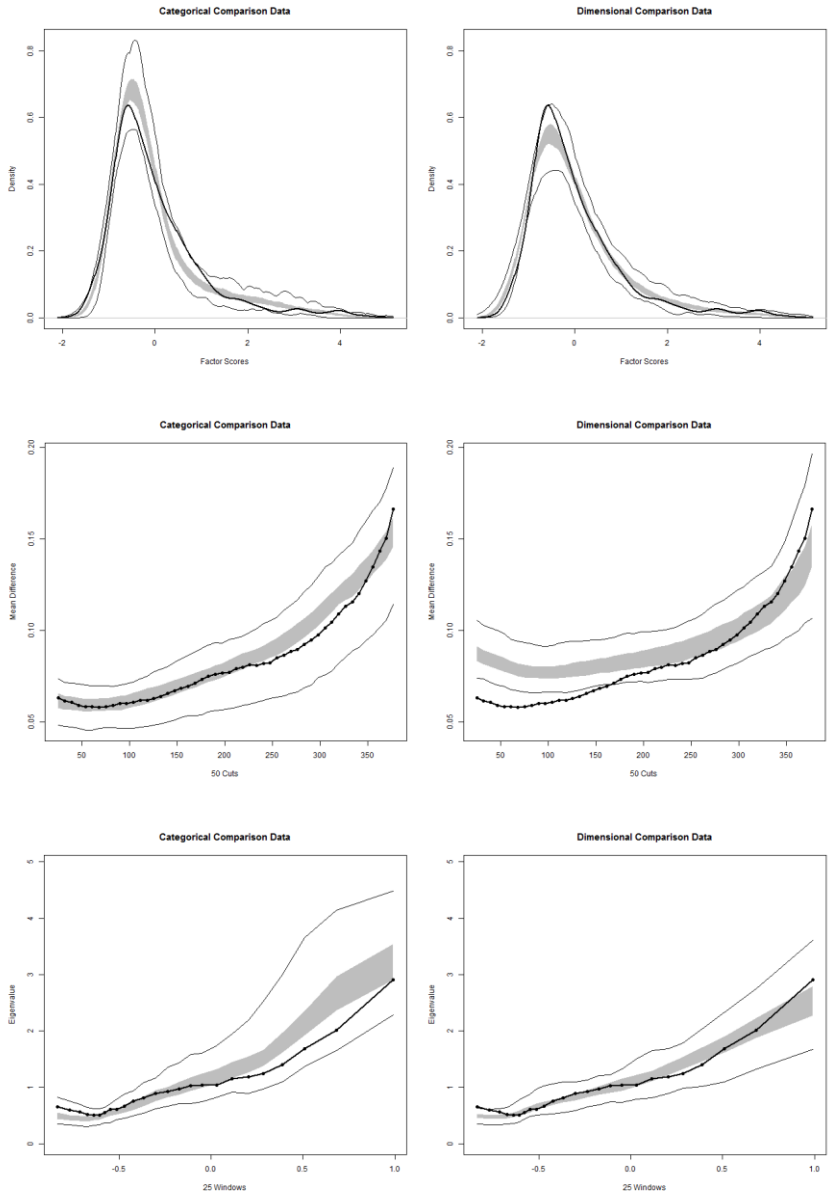

Figure A-7. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes in the RTC 1: Slide dataset.

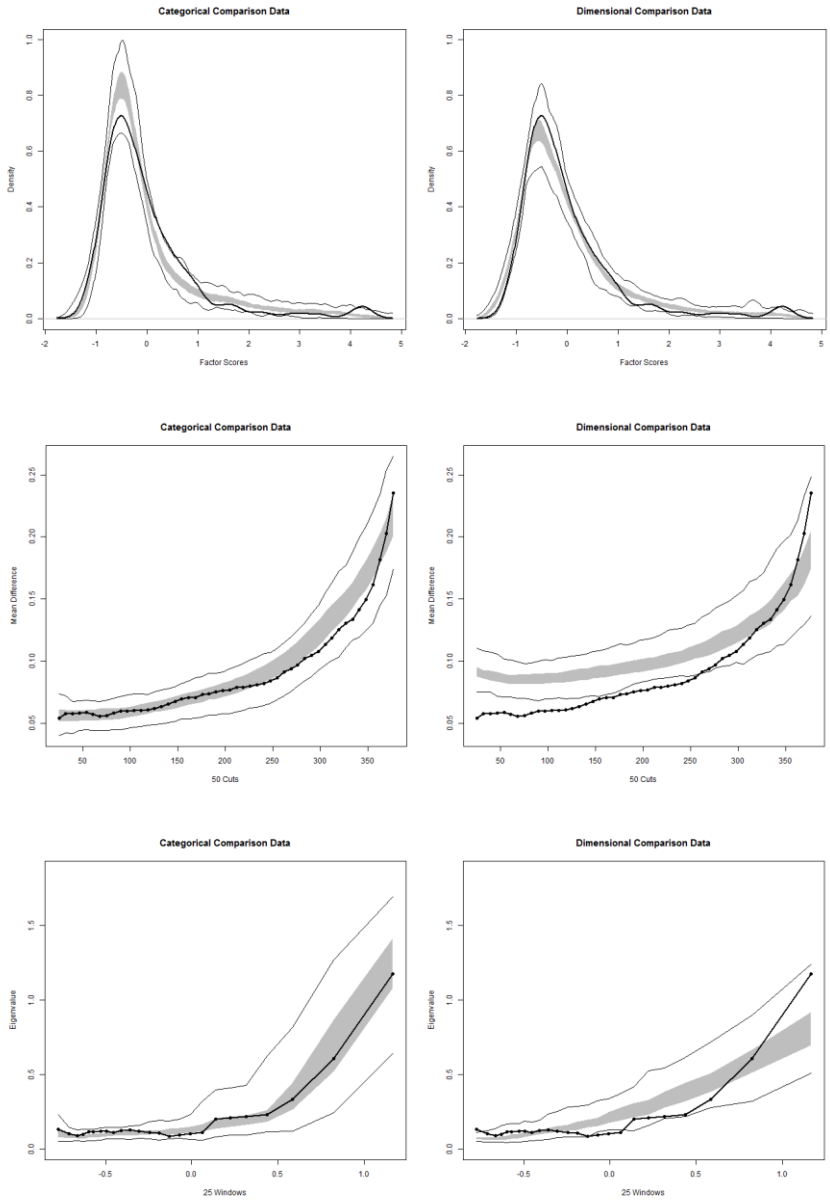

Figure A-8. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the RTC 1: Slide dataset.

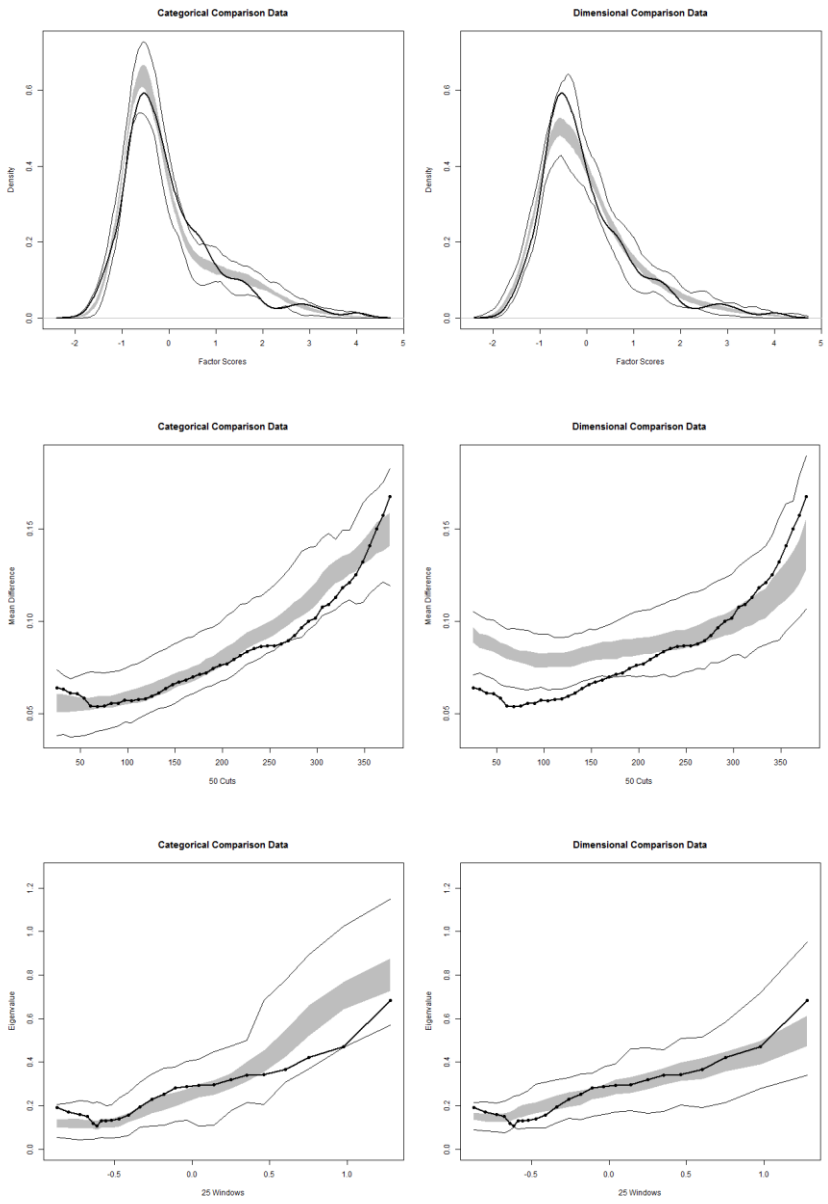

Figure A-9. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the RTC 1: Slide dataset.

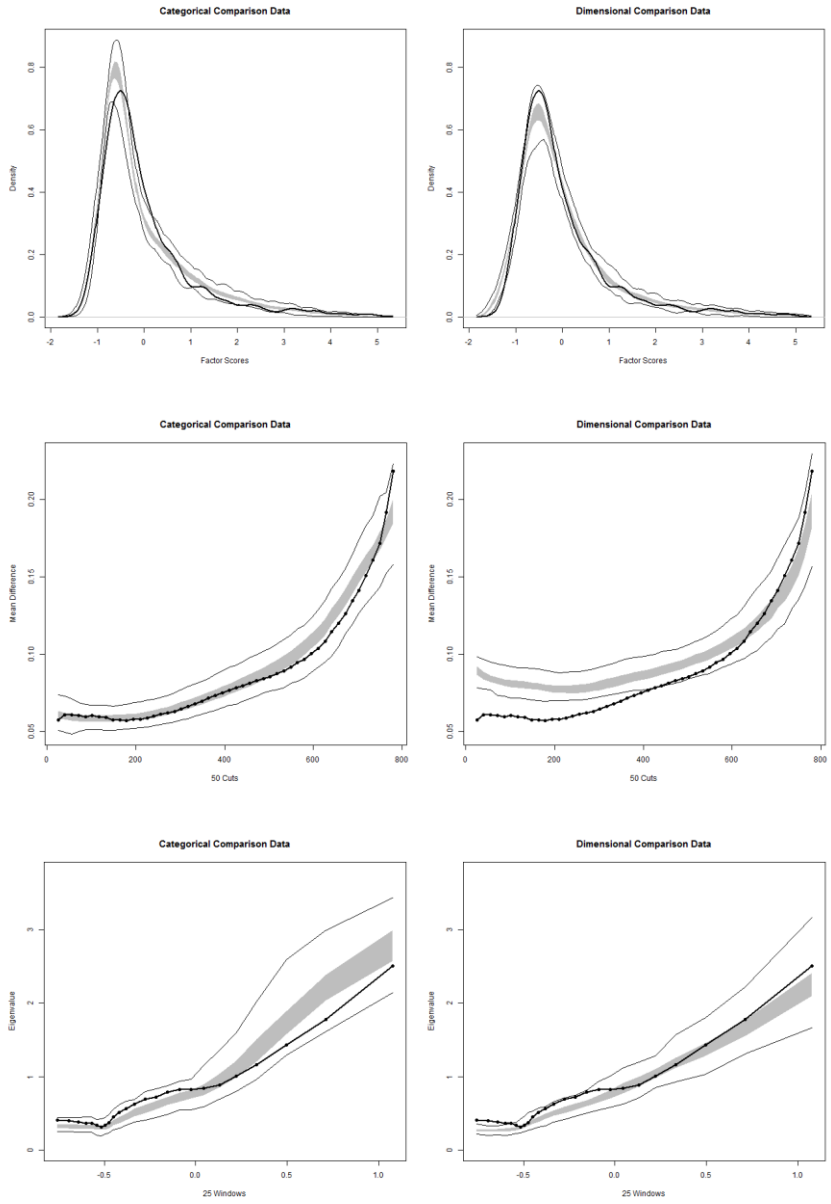

Figure A-10. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes in the RTC 2 dataset.

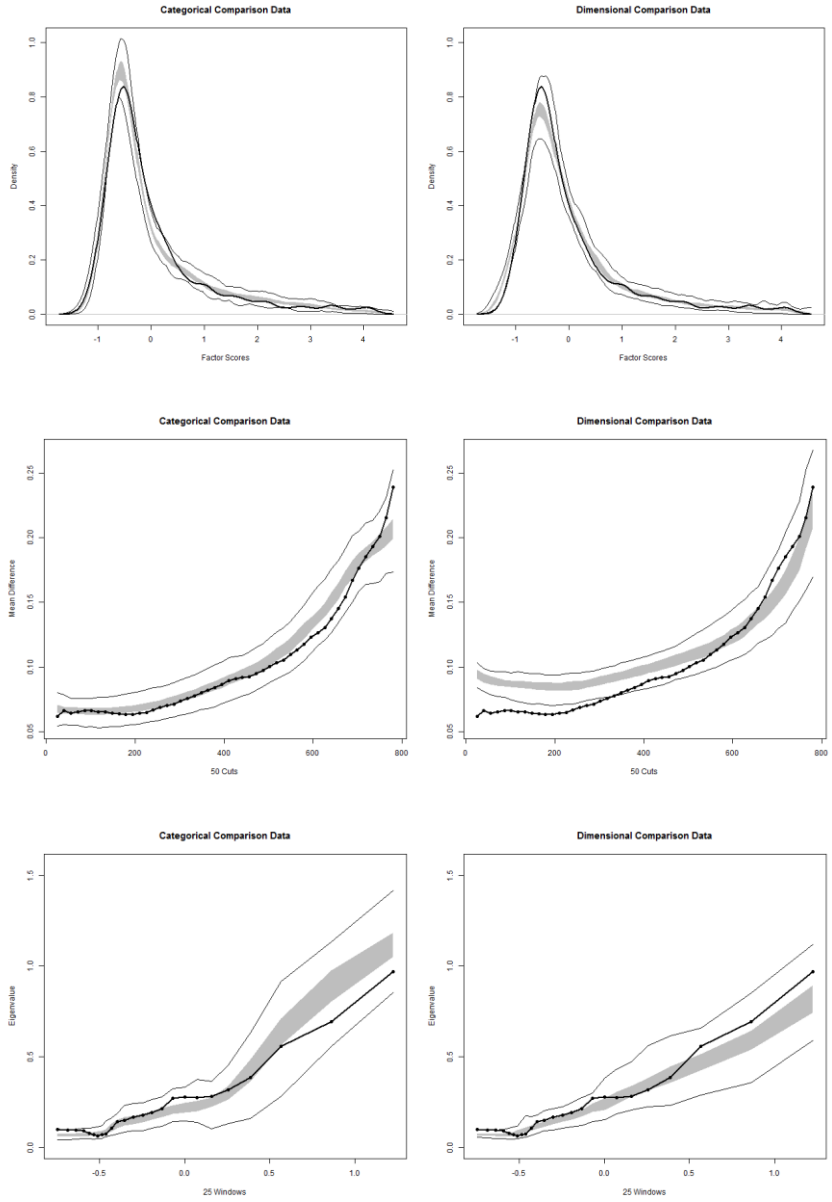

Figure A-11. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the RTC 2 dataset.

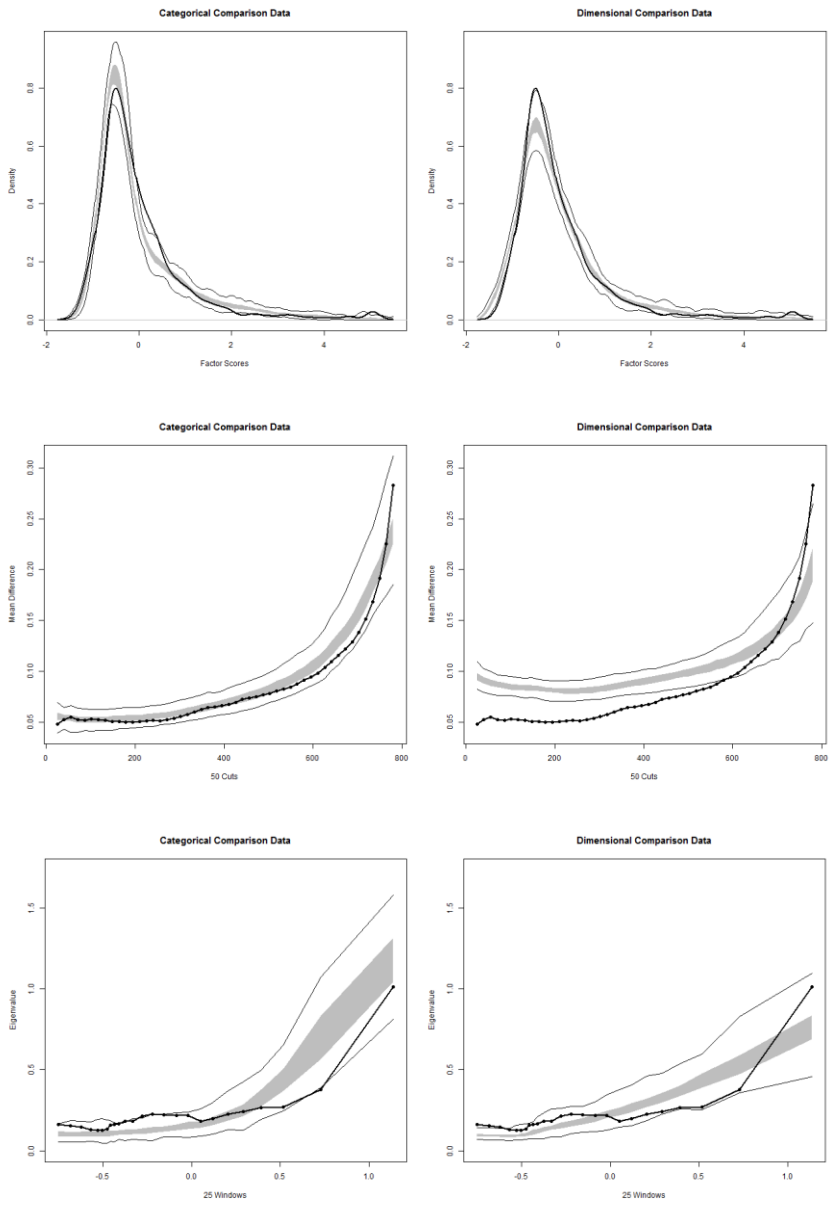

Figure A-12. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the RTC 2 dataset.

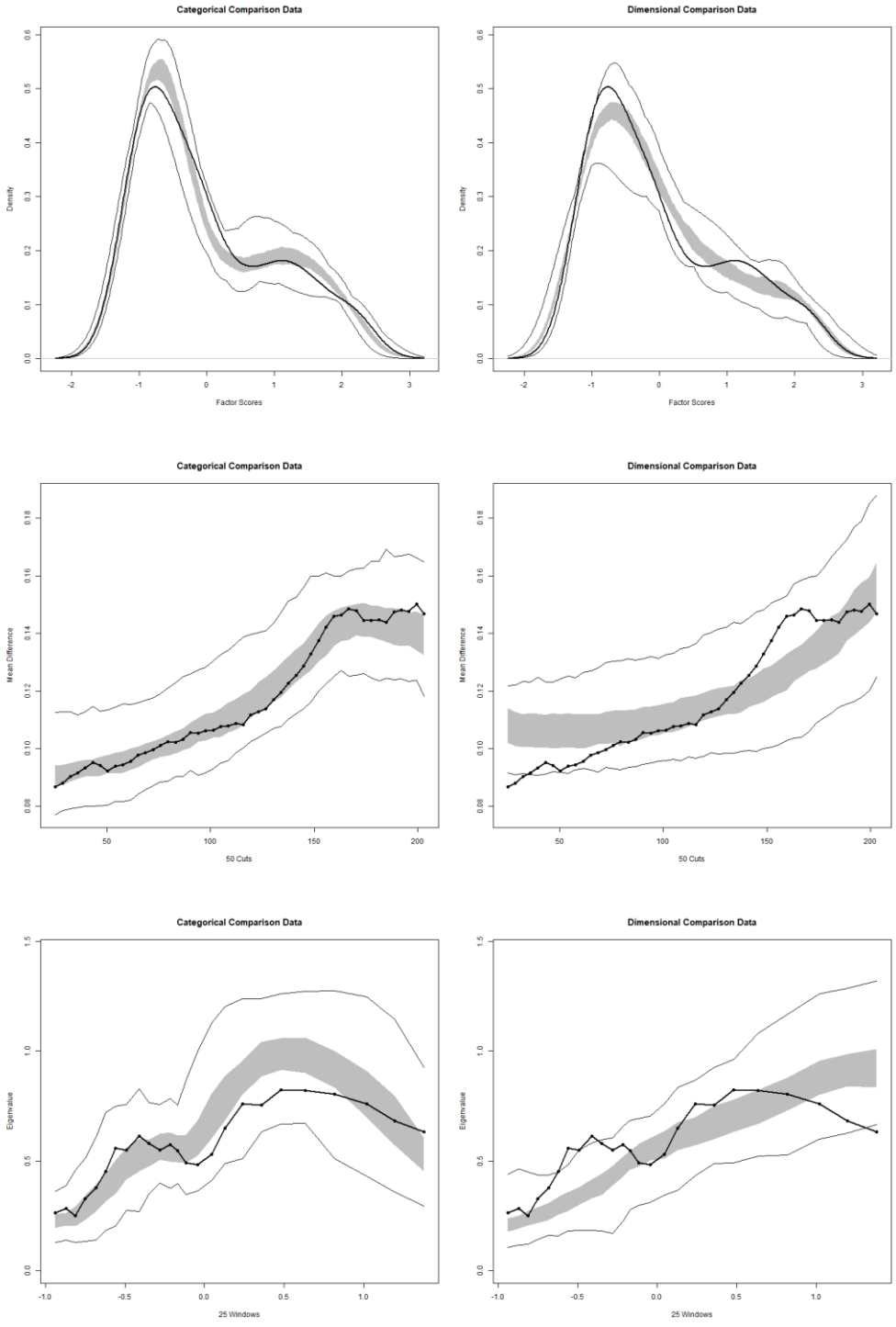

Figure A-13. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the combined audio datasets restricted to putative taxon members.

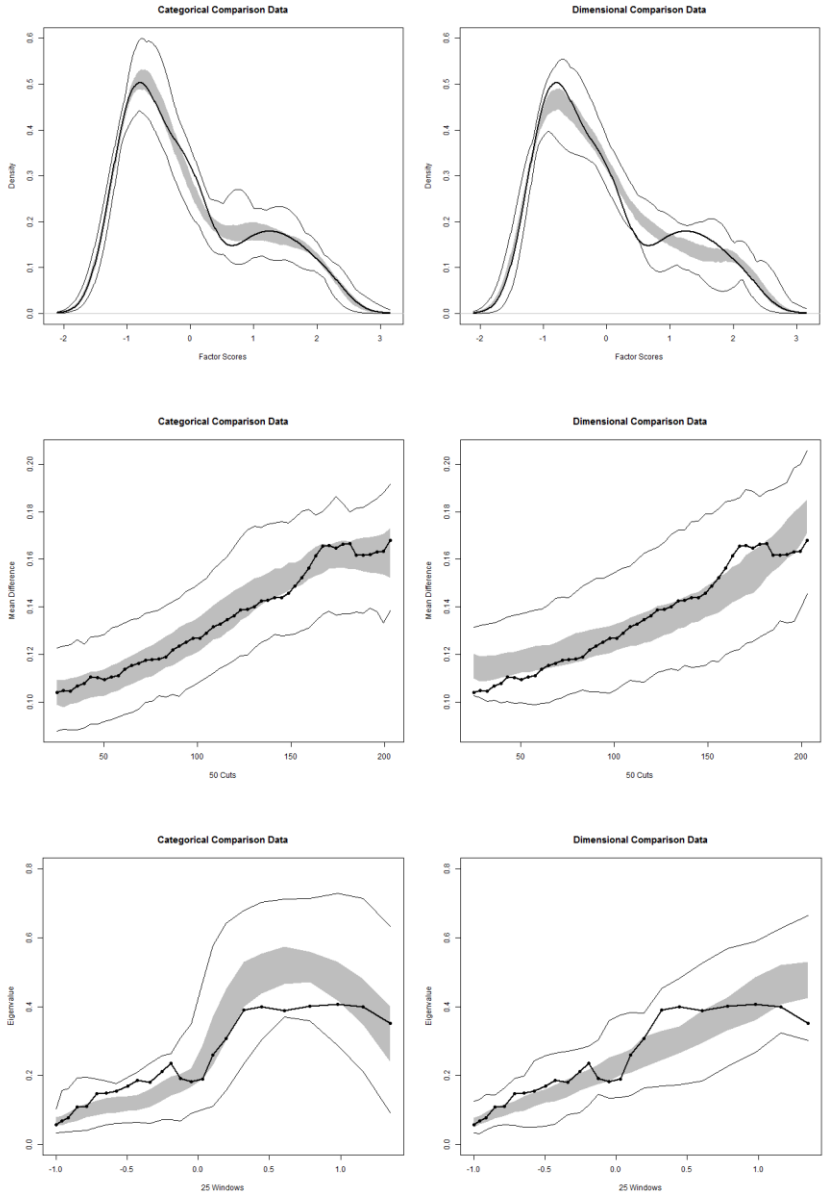

Figure A-14. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the combined audio datasets restricted to putative taxon members.

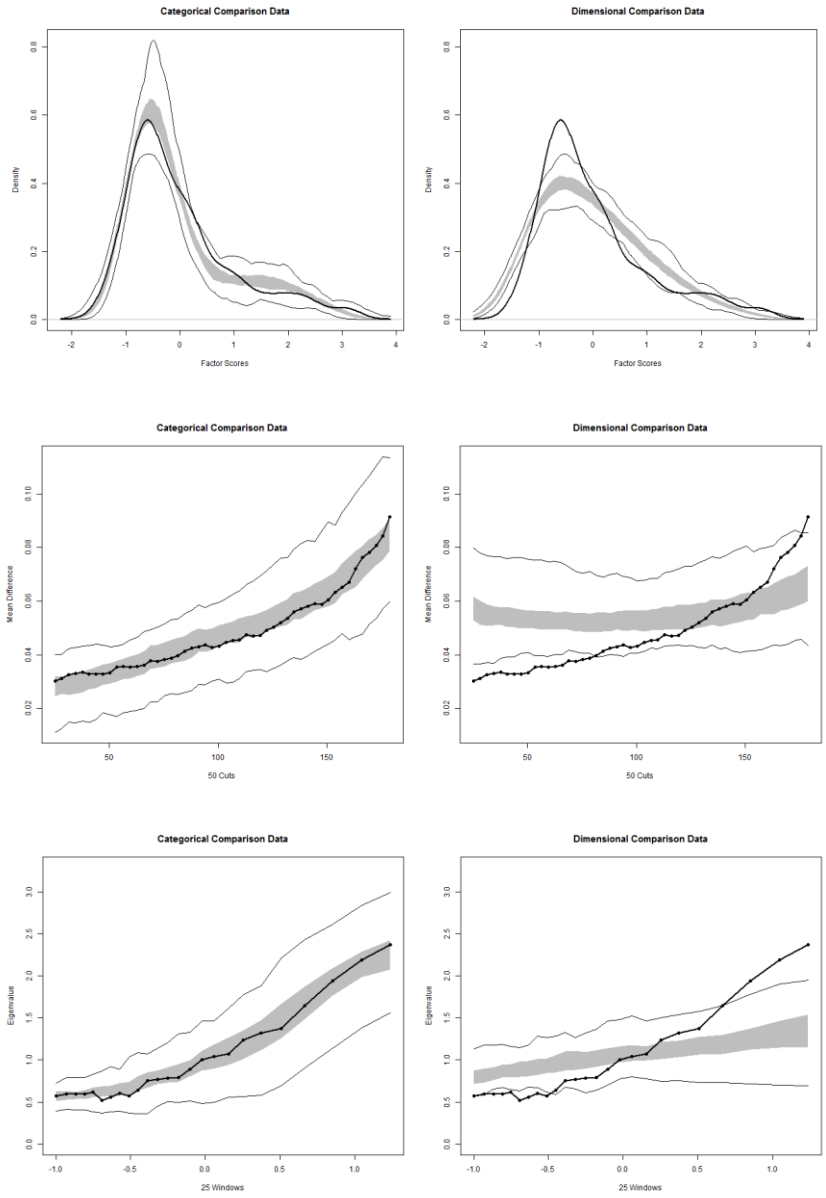

Figure A-15. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes in the RTC 2 dataset restricted to putative taxon members.

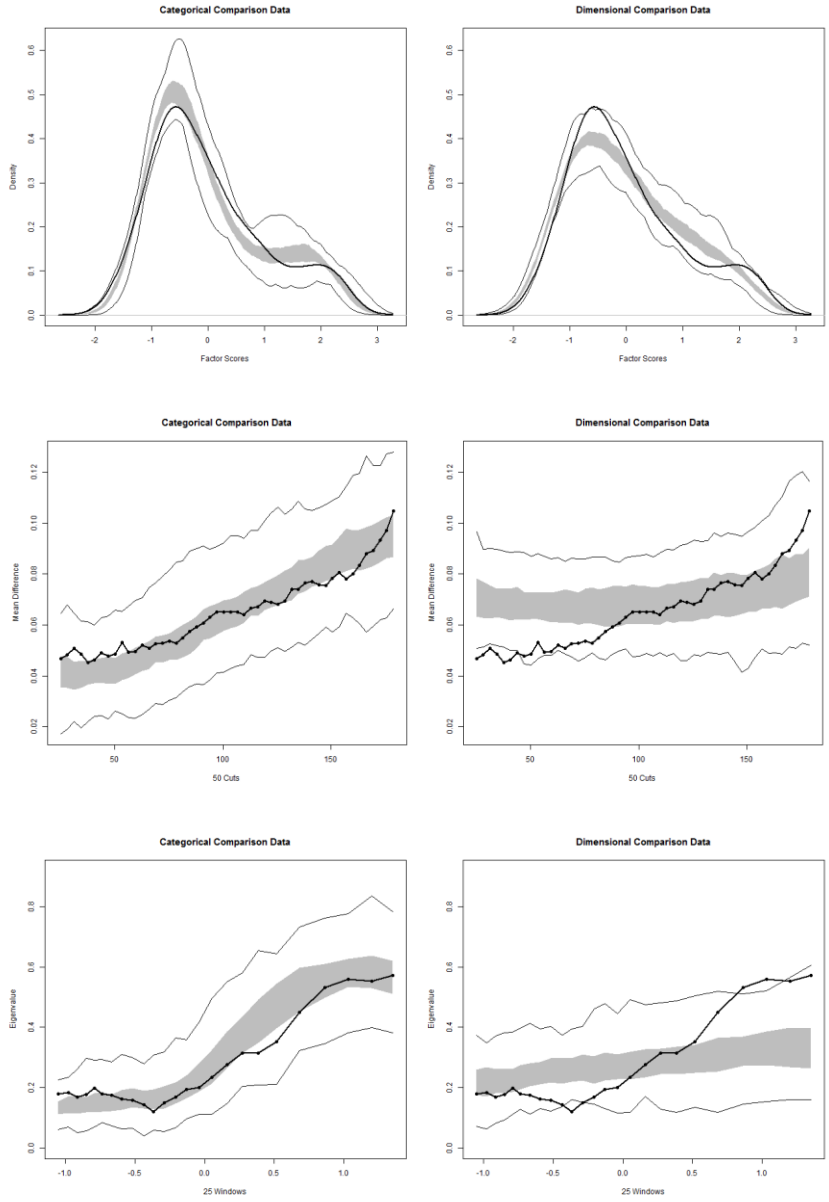

Figure A-16. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the RTC 2 dataset restricted to putative taxon members.

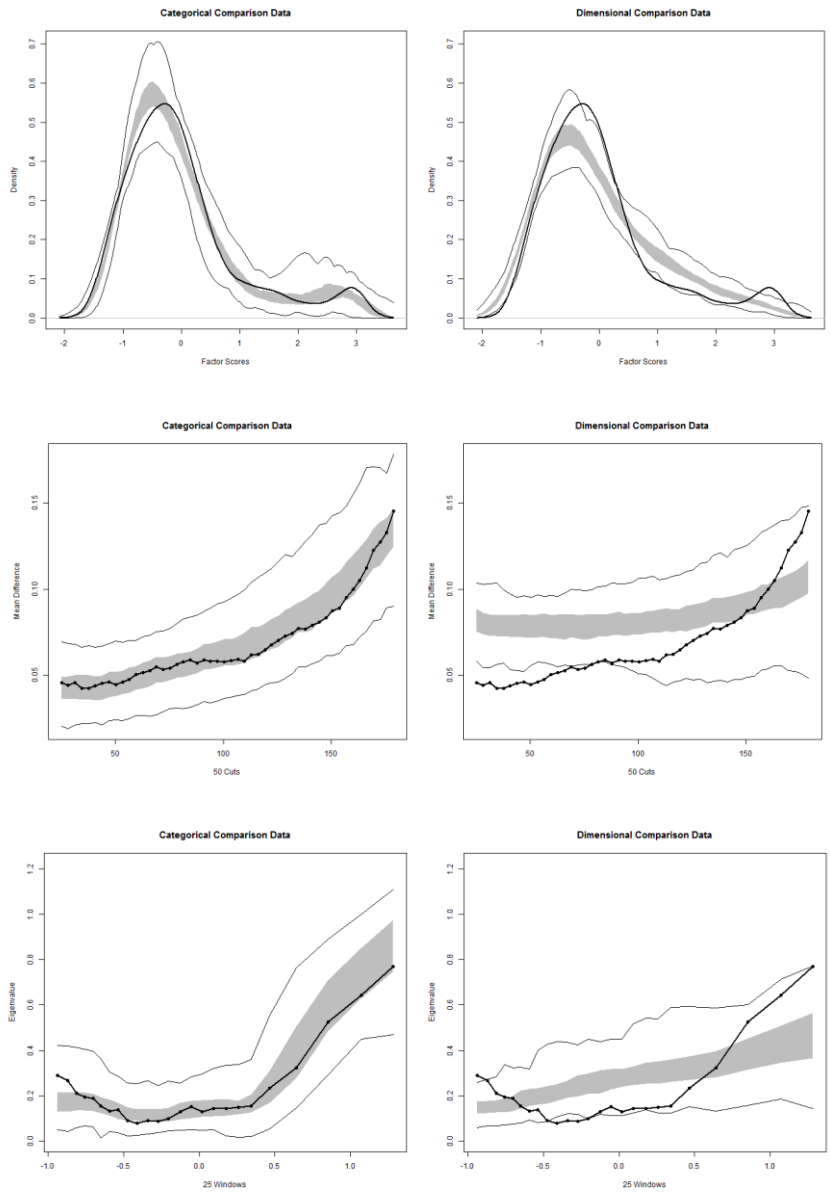

Figure A-17. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the RTC 2 dataset restricted to putative taxon members.

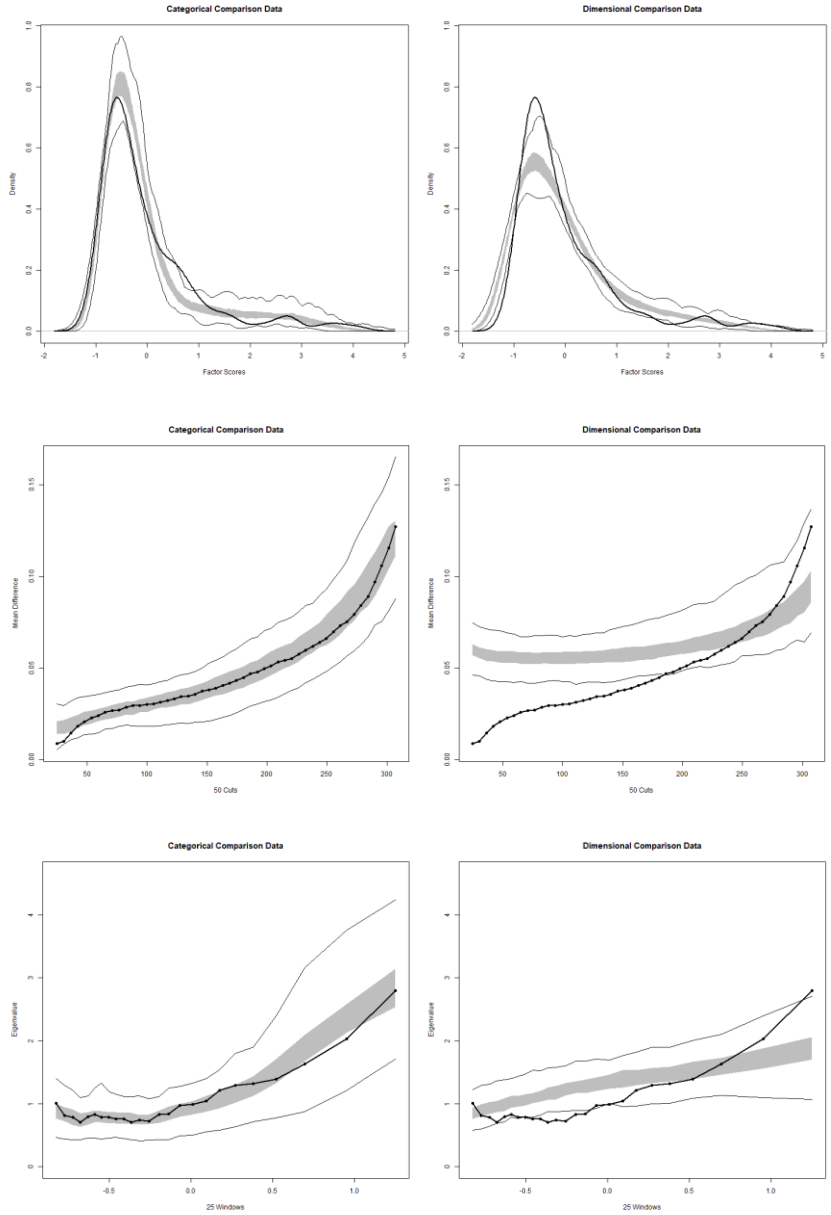

Figure A-18. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in both sexes in the combined slides dataset restricted to the putative taxon members.

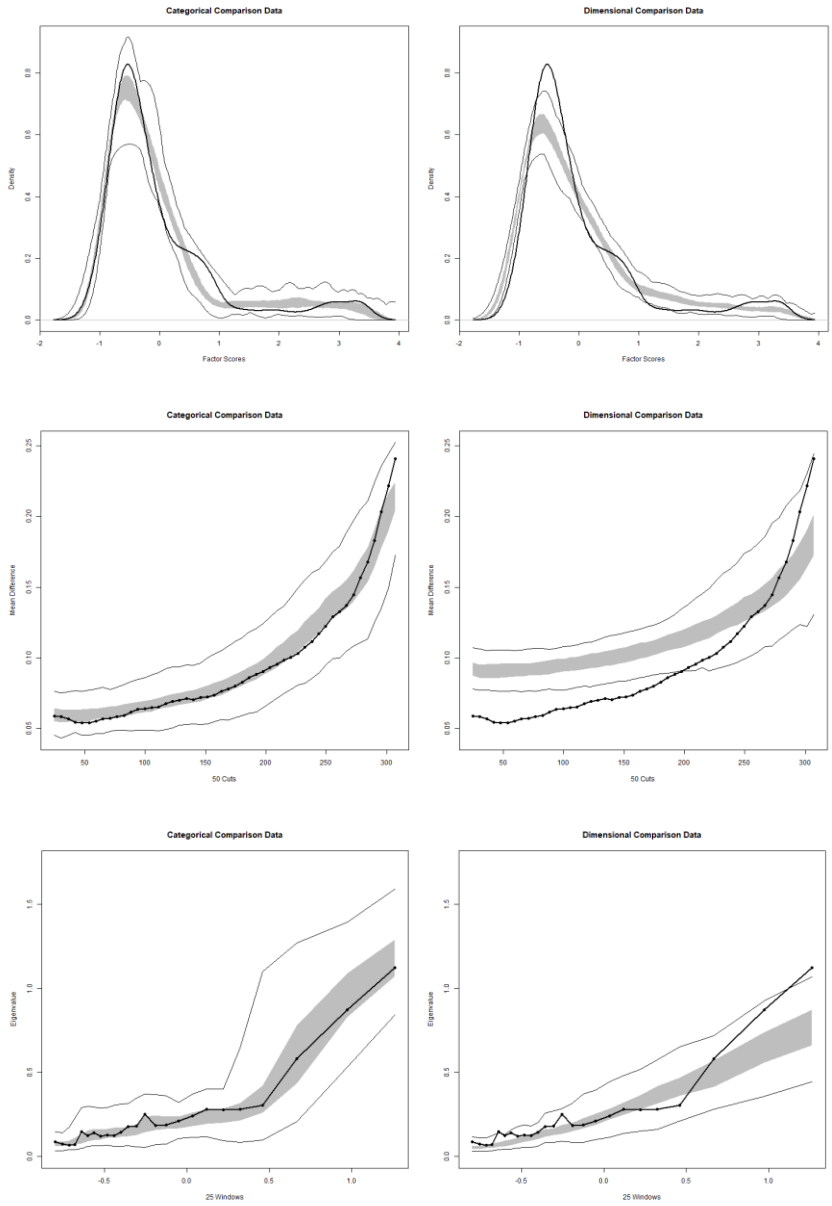

Figure A-19. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in female children in the combined slide datasets restricted to putative taxon members.

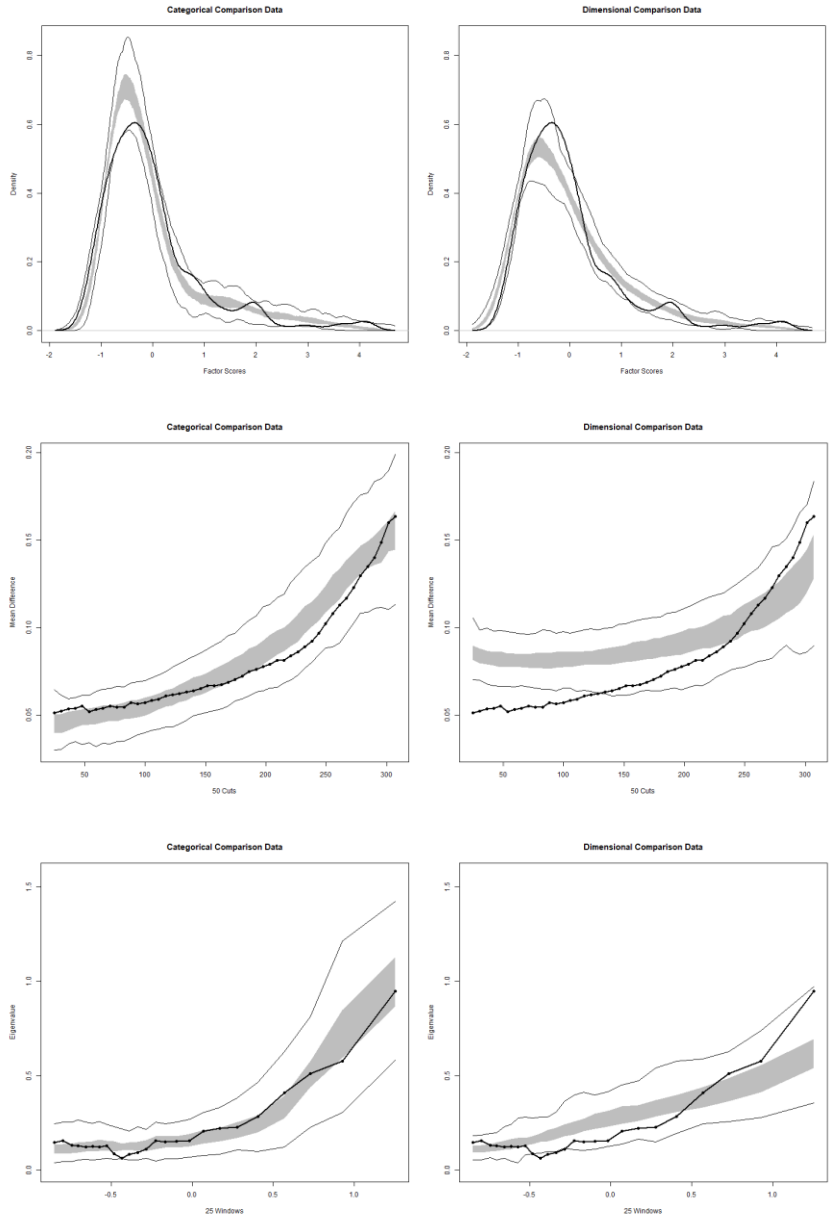

Figure A-20. L-Mode, MAMBAC, and MAXEIG graphs comparing research data to simulated categorical and dimensional data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data. Curves taken from the analysis of indicators of interest in male children in the combined slide datasets restricted to putative taxon members.

Appendix B: Supplemental Figures from Taxometric Analyses

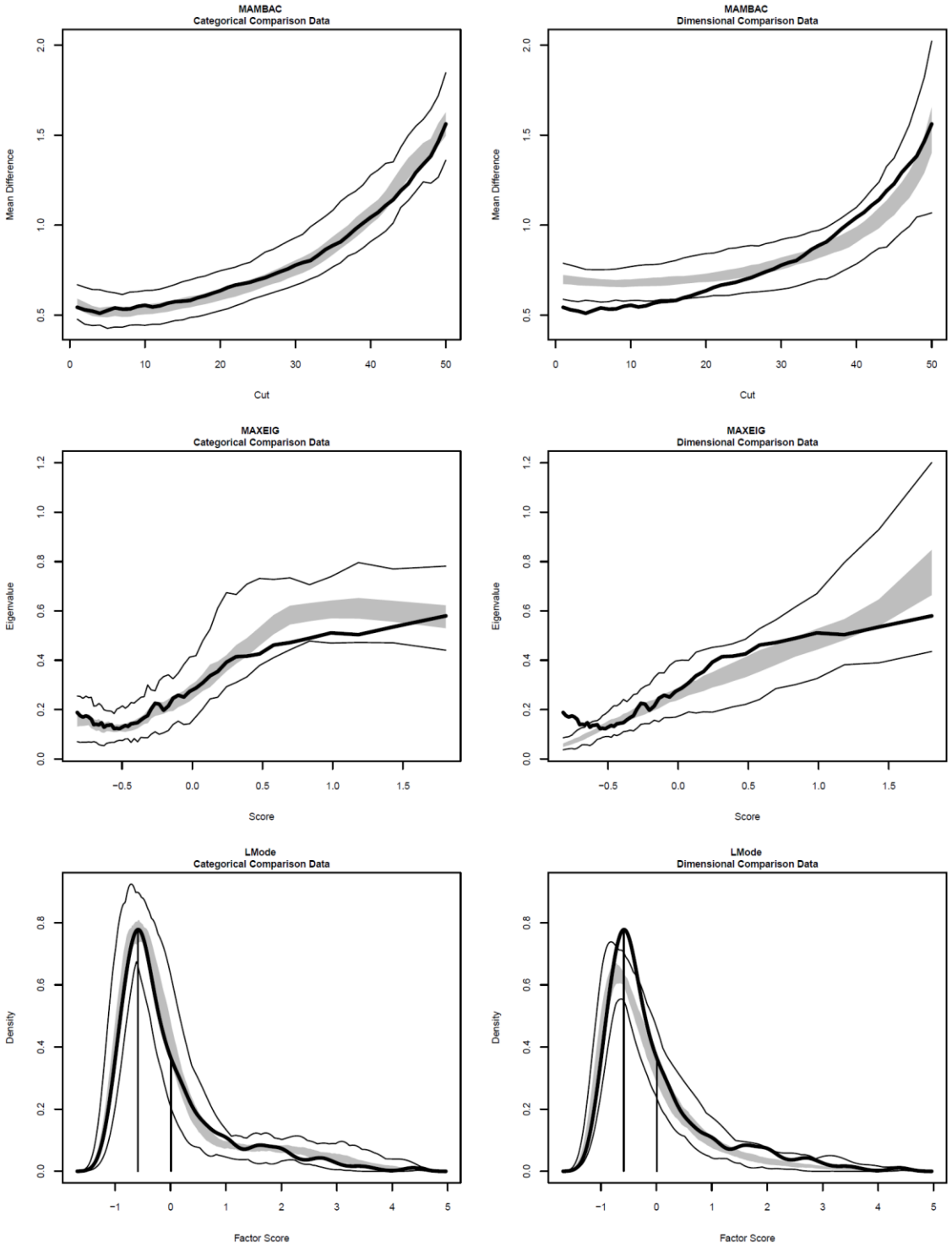

Figure B-1. MAMBAC, MAXEIG, and L-Mode curves for the Phillippe-Pinel dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data.

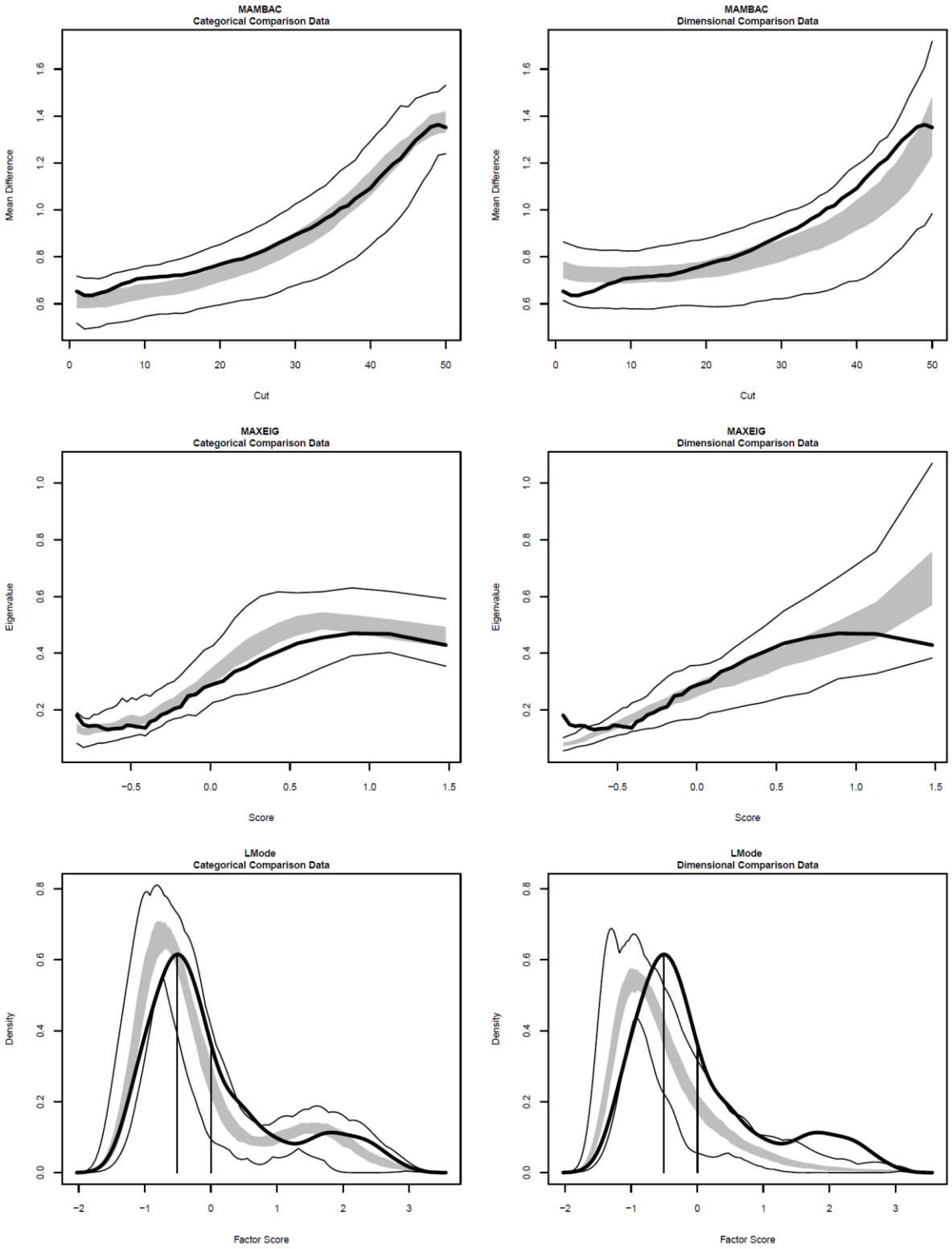

Figure B-2. MAMBAC, MAXEIG, and L-Mode curves for the RTC 1: Audio dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data.

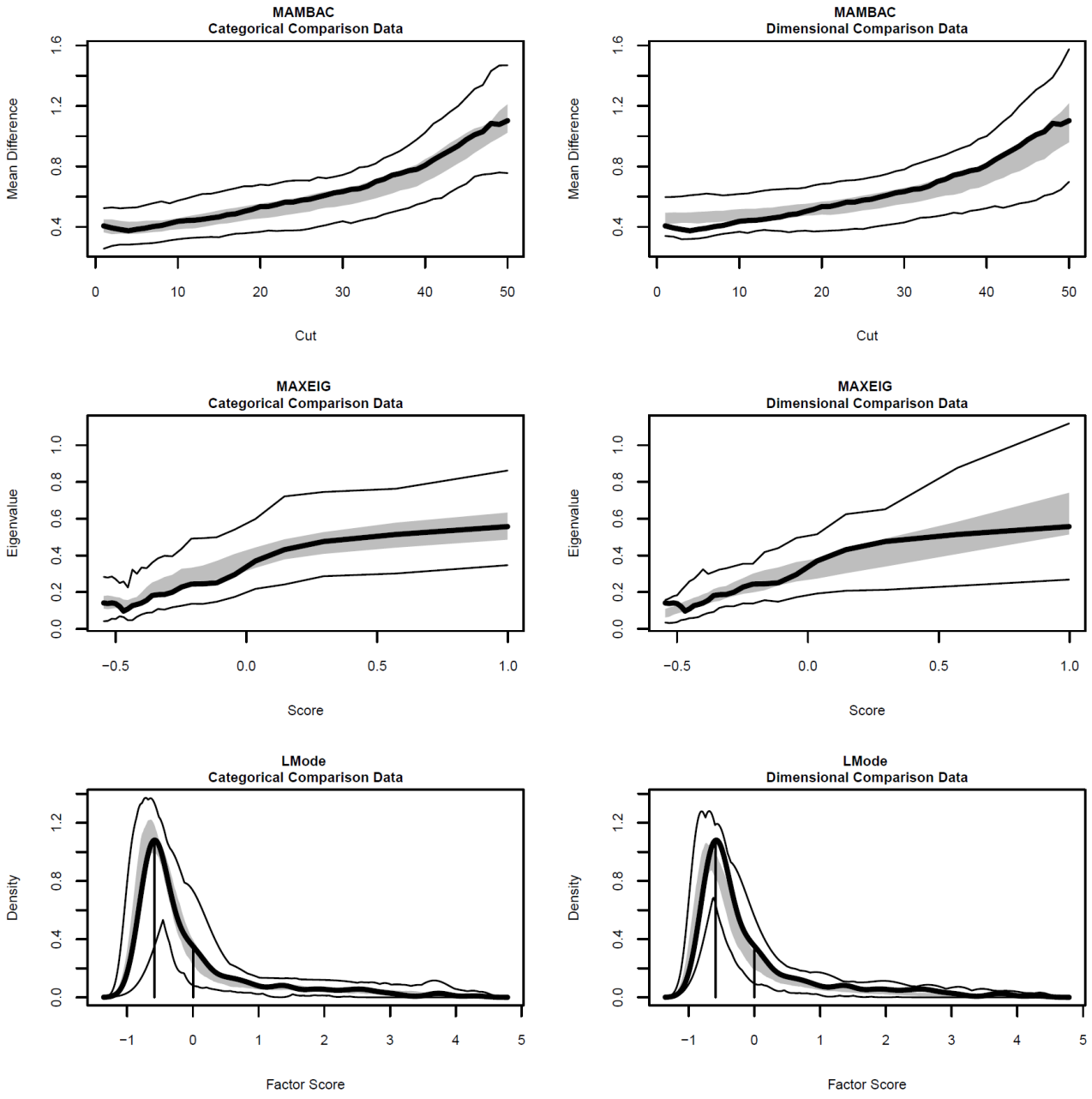

Figure B-3. MAMBAC, MAXEIG, and L-Mode curves for the RPC dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data.

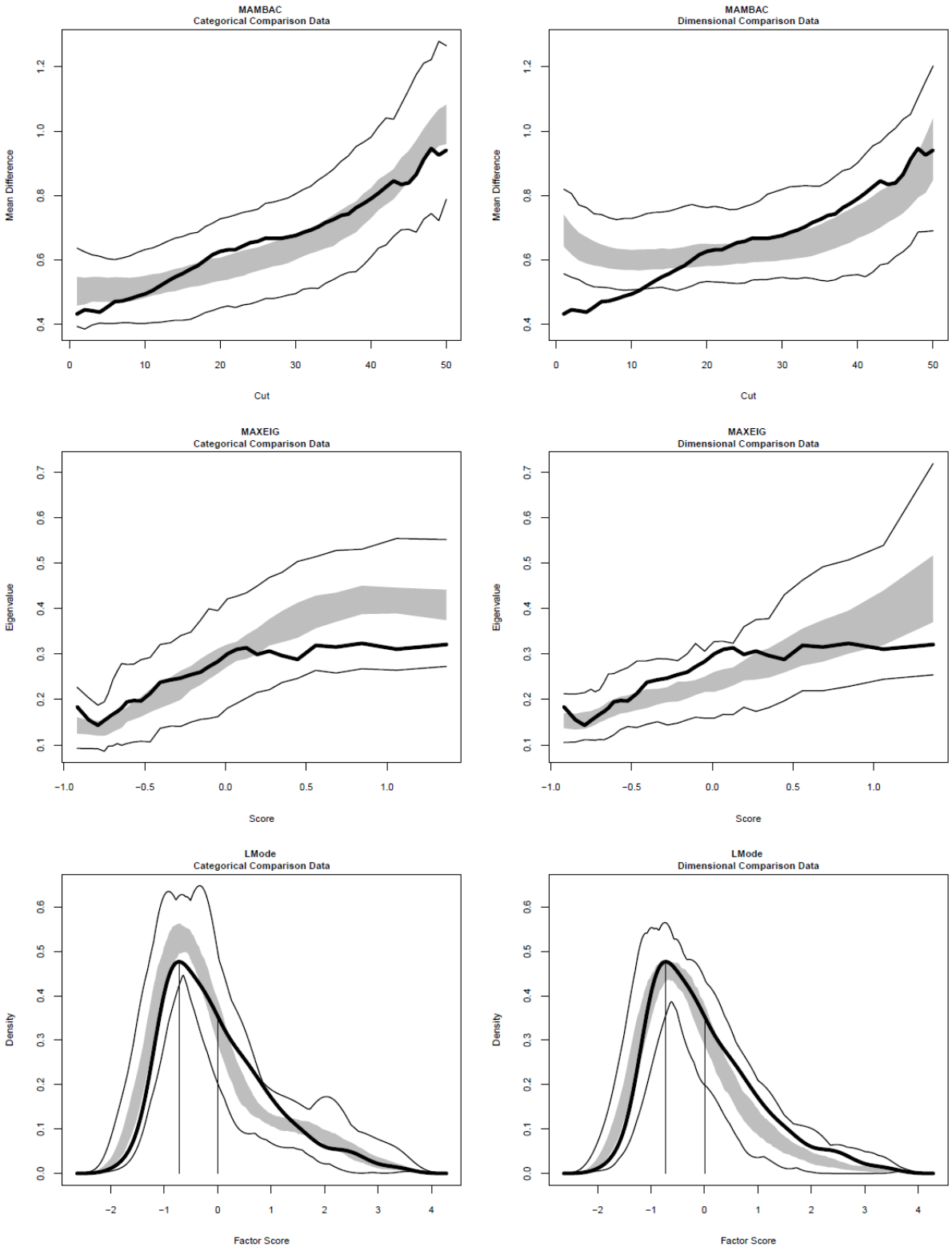

Figure B-4. MAMBAC, MAXEIG, and L-Mode curves for the RTC 1: Slide dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data.

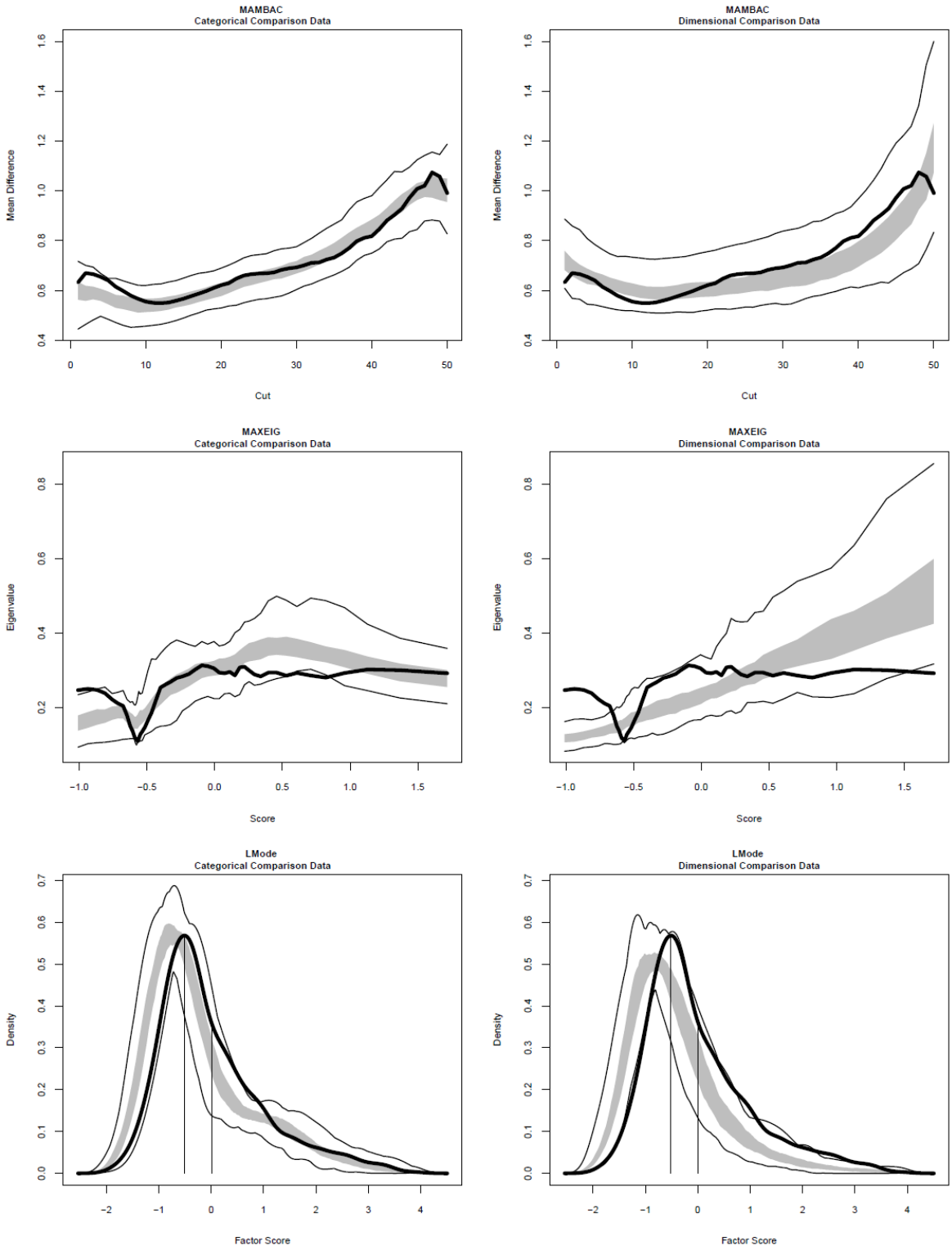

Figure B-5. MAMBAC, MAXEIG, and L-Mode curves for the RTC 2 dataset. The graphs compare research data to simulated categorical (left) and dimensional (right) data. The dark line represents the average data curve, the gray line represents the middle 50% of the simulated data and the light lines show the minimum and maximum values of the simulated data.

Appendix C: References for Studies Included in the Meta-Analysis

C.1 References for Studies Included in Meta-Analysis

Group-based studies

Aldridge, N. C. (1999). *Evaluating treatment for sex offenders: a pretest-posttest and follow-up study* (Doctoral dissertation). University of Georgia.

Bédard, G. (2009). *Un programme de traitement cognitif-béhavioral offert en communauté à des agresseurs sexuels: l'efficacité évaluée selon le type de victims* (Doctoral dissertation). Université de Montreal.

Bradford, J. M., & Pawlak, A. (1993). Effects of cyproterone acetate on sexual arousal patterns of pedophiles. *Archives of Sexual Behavior, 22,* 629-641. doi:10.1007/BF01543305

Clift, R. J., Rajlic, G., & Gretton, H. M. (2009). Discriminative and predictive validity of the penile plethysmograph in adolescent sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 21,* 335–362. doi:10.1177/1079063209338491

Cooper, A. J., Sandhu, S., Losztyn, S., & Cernovsky, Z. (1992). A double-blind placebo-controlled trial of medroxyprogesterone acetate and cyproterone acetate with seven pedophiles. *The Canadian Journal of Psychiatry/La Revue canadienne de psychiatrie, 37,* 687–693. doi:10.1177/070674379203701003

Crolley, J., Roys, D., Thyer, B. A., & Bordnick, P. S. (1998). Evaluating outpatient behavior therapy of sex offenders: A pretest-posttest study. *Behavior Modification, 22,* 485–501. doi:10.1177/01454455980224003

Dohrmann, M. (2009). *Treatment effects of EDMR on risk to re-offend by sexual offenders traumatized as children* (Doctoral dissertation). Colorado School of Professional Psychology.

Gray, S. R. (1995). A comparison of verbal satiation and minimal arousal conditioning to reduce deviant arousal in the laboratory. *Sexual Abuse: A Journal of Research and Treatment, 7,* 143–153. doi:10.1177/107906329500700204

Hunter, J. A., & Goodwin, D. W. (1992). The clinical utility of satiation therapy with juvenile sexual offenders: Variations and efficacy. *Sexual Abuse: A Journal of Research and Treatment, 5,* 71–80. doi:10.1177/107906329200500201

Hunter Jr, J. A., & Santos, D. R. (1990). The use of specialized cognitive-behavioral therapies in the treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 34,* 239–247. doi:10.1177/0306624X9003400307

Johnston, P. (1995). *Assessment and modification of sexual preference in child molesters* (Doctoral dissertation). University of Canterbury.

Johnston, P., Hudson, S. M., & Marshall, W. L. (1992). The effects of masturbatory reconditioning with nonfamilial child molesters. *Behaviour Research and Therapy, 30,* 559–561. doi:10.1016/0005-7967(92)90043-G

Jones, D. T. (2014). An evaluation of the use of phallometric assessment for men incarcerated for sexually offending against children in New Zealand: Past results and future directions (Doctoral dissertation). University of Waikato.

Kaplan, M. S., Morales, M., & Becker, J. V. (1993). The impact of verbal satiation on adolescent sex offenders: A preliminary report. *Journal of Child Sexual Abuse, 2,* 81–88. doi:10.1300/J070v02n03_06

Lang, L. H. (1990). *The effects of masturbatory satiation on sexual arousal to children* (Doctoral dissertation). University of Texans at Arlington.

Marques, J., Nelson, C., West, M. A., & Day, D. M. (1994). The relationship between treatment goals and recidivism among child molesters. *Behaviour Research and Therapy, 32,* 577–588. doi:10.1016/0005-7967(94)90148-1

Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & Van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment, 17,* 79–107. doi:10.1177/107906320501700108

Marshall, W. L., & Barbaree, H. E. (1988). The long-term evaluation of a behavioral treatment program for child molesters. *Behaviour Research and Therapy, 26,* 499–511. doi:10.1016/0005-7967(88)90146-5

Marshall, W. L. (1997). The relationship between self-esteem and deviant sexual arousal in nonfamilial child molesters. *Behavior Modification, 21,* 86–96. doi:10.1177/01454455970211005

Quinsey, V. L., Bergersen, S. G., & Steinman, C. M. (1976). Changes in physiological and verbal responses of child molesters during aversion therapy. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 8,* 202–212. doi:10.1037/h0081948

Quinsey, V. L., Chaplin, T. C., & Carrigan, W. F. (1980). Biofeedback and signaled punishment in the modification of inappropriate sexual age preferences. *Behavior Therapy, 11,* 567–576. doi:10.1016/S0005-7894(80)80072-4

Ricci, R. J., Clayton, C. A., & Shapiro, F. (2006). Some effects of EMDR on previously abused child molesters: Theoretical reviews and preliminary findings. *The Journal of Forensic Psychiatry & Psychology, 17,* 538–562. doi:10.1080/14789940601070431

Rice, M. E., Quinsey, V. L., & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum-security psychiatric institution. *Journal of Consulting and Clinical Psychology, 59,* 381–386.

Schober, J. M., Kuhn, P. J., Kovacs, P. G., Earle, J. H., Byrne, P. M., & Fries, R. A. (2005). Leuprolide acetate suppresses pedophilic urges and arousability. *Archives of Sexual Behavior, 34,* 691–705. doi:10.1007/s10508-005-7929-2

Weinrott, M. R., Riggan, M., & Frothingham, S. (1997). Reducing deviant arousal in juvenile sex offenders using vicarious sensitization. *Journal of Interpersonal Violence, 12,* 704–728. doi:10.1177/088626097012005007

Single case designs

Alford, G. S., Morin, C., Atkins, M., & Schoen, L. (1987). Masturbatory extinction of deviant sexual arousal: A case study. *Behavior Therapy, 18,* 265–271. doi:10.1016/S0005-7894(87)80020-5

Earls, C. M., & Castonguay, L. G. (1989). The evaluation of olfactory aversion for a bisexual pedophile with a single-case multiple baseline design. *Behavior Therapy, 20,* 137–146. doi:10.1016/S0005-7894(89)80125-X

Foote, W. E., & Laws, D. R. (1981). A daily alternation procedure for orgasmic reconditioning with a pedophile. *Journal of Behavior Therapy and Experimental Psychiatry, 12,* 267–273. doi:10.1016/0005-7916(81)90057-4

Harbert, T. L., Barlow, D. H., Hersen, M., & Austin, J. B. (1974). Measurement and modification of incestuous behavior: A case study. *Psychological Reports, 34,* 79–86. doi:10.2466/pr0.1974.34.1.79

Kremsdorf, R. B., Holmen, M. L., & Laws, D. R. (1980). Orgasmic reconditioning without deviant imagery: A case report with a pedophile. *Behaviour Research and Therapy, 18,* 203–207. doi:10.1016/0005-7967(80)90037-6

Laws, D. R. (1980). Treatment of bisexual pedophilia by a biofeedback-assisted self-control procedure. *Behaviour Research and Therapy, 18,* 207–211. doi:10.1016/0005-7967(80)90038-8

Laws, D. R. (2001). Olfactory aversion: Notes on procedure, with speculations on its mechanism of effect. *Sexual Abuse: A Journal of Research and Treatment, 13,* 275–287. doi:10.1177/107906320101300405

Lee, F. C. (1988). *Massed conditioning with valeric acid in the reduction of penile circumference of heterosexual pedophiles* (Doctoral dissertation). University of Denver.

Levin, S. M., Barry, S. M., Gambaro, S., Wolfinsohn, L., & Smith, A. (1977). Variations of covert sensitization in the treatment of pedophilic behavior: A case study. *Journal of Consulting and Clinical Psychology, 45,* 896–907. doi:10.1037/0022-006X.45.5.896

Marshall, W. L. (1979). Satiation therapy: A procedure for reducing deviant sexual arousal. *Journal of Applied Behavior Analysis, 12,* 377–389. doi:10.1901/jaba.1979.12-377

Marshall, W. L. (2006). Olfactory aversion and directed masturbation in the modification of deviant preferences: A case study of a child molester. *Clinical Case Studies, 5,* 3–14. doi:0.1177/1534650103259754

Marshall, W. L. (2007). Covert association: A case demonstration with a child molester. *Clinical Case Studies, 6,* 218–231. doi:10.1177/1534650105280329

Marshall, W. L., & Barbaree, H. E. (1978). The reduction of deviant arousal: Satiation treatment for sexual aggressors. *Correctional Psychologist, 5,* 294–303. doi:10.1177/009385487800500402

Plaud, J. J., & Gaither, G. A. (1997). A clinical investigation of the possible effects of long-term habituation of sexual arousal in assisted covert sensitization. *Journal of Behavior Therapy and Experimental Psychiatry, 28,* 281–290. doi:10.1016/S0005-7916(97)00026-8

Rea, J. A., Williams, D., Saunders, K. J., Dixon, M., Wright, K., & Spradlin, J. E. (2003). Covert sensitization: A generalization analysis in the laboratory and natural environment through the use of a portable-penile plethysmograph. *The Behavior Analyst Today, 4,* 192–201. doi:10.1037/h0100119

Stava, L., Levin, S. M., & Schwanz, C. (1993). The role of aversion in covert sensitization treatment of pedophilia: A case report. *Journal of Child Sexual Abuse, 2,* 1-14. doi:10.1300/J070v02n03_01

VanDeventer, A. D., & Laws, D. R. (1978). Orgasmic reconditioning to redirect sexual arousal in pedophiles. *Behavior Therapy, 9,* 748–765. doi:10.1016/S0005-7894(78)80006-9

Wincze, J. P., Bansal, S., & Malamud, M. (1986). Effects of medroxyprogesterone acetate on subjective arousal, arousal to erotic stimulation, and nocturnal penile tumescence in male sex offenders. *Archives of Sexual Behavior, 15,* 293-305. doi:10.1007/BF01550364

C.2 Studies Included in the Test-Retest Meta-Analyses

Babchishin, K. M., Curry, S. D., Fedoroff, J. P., Bradford, J., & Seto, M. C. (2017). Inhibiting Sexual Arousal to Children: Correlates and Its Influence on the Validity of Penile Plethysmography. *Archives of Sexual Behavior, 46,* 671–684. doi:10.1007/s10508-017-0952-2

Becker, J. V., Hunter, J. A., Goodwin, D., Kaplan, M. S., & Martinez, D. (1992). Test-retest reliability of audio-taped phallometric stimuli with adolescent sexual offenders. *Annals of Sex Research, 5,* 45–51. doi:10.1007/BF00849876

Eccles, A., Marshall, W. L., & Barbaree, H. E. (1988). The vulnerability of erectile measures to repeated assessments. *Behaviour Research and Therapy, 26,* 179–183. doi:10.1016/0005-7967(88)90117-9

Fernandez, Y. M. (1998). *Phallometric testing with sexual offenders against female victims: An examination of reliability and validity issues* (Doctoral dissertation). Queen's University.

Gray, S. R. (1995). A comparison of verbal satiation and minimal arousal conditioning to reduce deviant arousal in the laboratory. *Sexual Abuse: A Journal of Research and Treatment, 7,* 143–153. doi:10.1177/107906329500700204

Rice, M. E., Harris, G. T., Lang, C., & Chaplin, T. C. (2008). Sexual preferences and recidivism of sex offenders with mental retardation. *Sexual Abuse: A Journal of Research and Treatment, 20,* 409–425. doi:10.1177/1079063208324662

Wormith, J. S. (1986). Assessing deviant sexual arousal: Physiological and cognitive aspects. *Advances in Behaviour Research and Therapy, 8,* 101–137. doi:10.1016/0146-6402(86)90001-9

C.3 Studies Included Constructing Natural History Benchmarks

Fernandez, Y. M. (1998). *Phallometric testing with sexual offenders against female victims: An examination of reliability and validity issues* (Doctoral dissertation). Queen's University.

Ricci, R. J., Clayton, C. A., & Shapiro, F. (2006). Some effects of EMDR on previously abused child molesters: Theoretical reviews and preliminary findings. *The Journal of Forensic Psychiatry & Psychology, 17,* 538–562. doi:10.1080/14789940601070431

Rice, M. E., Quinsey, V. L., & Harris, G. T. (1991). Sexual recidivism among child molesters released from a maximum-security psychiatric institution. *Journal of Consulting and Clinical Psychology, 59,* 381–386.

Weinrott, M. R., Riggan, M., & Frothingham, S. (1997). Reducing deviant arousal in juvenile sex offenders using vicarious sensitization. *Journal of Interpersonal Violence, 12,* 704–728. doi:10.1177/088626097012005007

Wormith, J. S. (1986). Assessing deviant sexual arousal: Physiological and cognitive aspects. *Advances in Behaviour Research and Therapy, 8,* 101–137. doi:10.1016/0146-6402(86)90001-9

C.4 Studies Included Constructing Normative Comparisons

Barsetti, I., Earls, C. M., Lalumière, M. L., & Bélanger, N. (1998). The differentiation of intrafamilial and extrafamilial heterosexual child molesters. *Journal of Interpersonal Violence, 13,* 275–286.

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, 335-–350.

Byrne, P. M. (2001). *The reliability and validity of less explicit audio and "clothed" visual penile plethysmograph stimuli with child molesters and nonoffenders* (Unpublished doctoral dissertation). University of Utah.

Grossman, L. S., Cavanaugh, J. L., & Haywood, T. W. (1992). Deviant sexual responsiveness on penile plethysmography using visual stimuli: Alleged child molesters vs. normal control subjects. *Journal of Nervous and Mental Disease, 180,* 207–208.

Harris, G. T., Rice, M. E., Quinsey, V. L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behaviour Research and Therapy, 34,* 389–394. doi:10.1016/0005-7967(95)00070-4

Harris, G. T., Rice, M. E., Quinsey, V. L., Chaplin, T. C., & Earls, C. (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment, 4,* 502–511.

Howes, R. J. (2014). *Phallometric test data from a sample of federally incarcerated Canadian offenders*. Unpublished manuscript.

Marshall, W. L., & Barbaree, H. E. (1988). The long-term evaluation of a behavioral treatment program for child molesters. *Behavioral Research and Therapy, 26*, 499–511. doi:10.1016/0005-7967(88)90146-5

Goyette, M. (2012). *Évaluation des intérêts sexuels auprès d'agresseurs sexuels d'enfants par la vidéo-oculographie et la présentation de stimuli générés par ordinateur* [Assessment of sexual interests with sexual aggressors against children by video-oculography and presentation of computer-generated stimuli] (Unpublished doctoral dissertation). Université de Montréal.

Looman, J. (2011). *Phallometric test data from a sample of federally incarcerated Canadian sexual offenders*. Unpublished manuscript.

McCoy, M.E. (1997). *Predictors of recidivism in a population of Canadian sex offenders: Psychological, physiological, and offence factors* (Unpublished doctoral dissertation). University of Ottawa.

Renaud, R., Chartier, R., Rouleau, J. L., Proulx, J., Goyette, M., Trottier, D., . . . Bouchard, S. (2013). Using immersive virtual reality and ecological psychology to probe into child molesters' phenomenology. *Journal of Sexual Aggression*, *19*, 102–120. doi:10.1080/13552600.2011.617014

Appendix D: Benchmark Comparisons

Table D-1

Benchmark comparisons for overarching treatment intervention modalities

	df	λ	Non-central *t* (upper 5%/1%)	g_{cv} (α = .05)	g_{cv} (α = .01)	$g_{treatment}$	Sig.
Behavioural Treatments							
Pedohebephilic interest	407	6.322	8.036/8.760	0.397	0.434	0.612	**
with single case designs	431	6.505	8.219/8.942	0.395	0.430	0.657	**
Pedophilic interest	55	2.342	4.126/4.928	0.551	0.659	0.778	**
with single case designs	63	2.504	4.280/5.071	0.535	0.634	0.840	**
Teleiophilic interest	74	--	--	--	--	–0.103	ns
Pharmacological Treatments							
Pedohebephilic interest	33	1.825	3.652/4.505	0.626	0.773	0.648	*
Comprehensive Treatments							
Pedohebephilic interest	586	7.583	9.291	0.383	--	0.202	ns
Pedophilic interest	427	--	--	--	--	0.122	ns
Teleiophilic interest	473	2.700	4.364	0.200	--	0.196	ns

Table D-2

Benchmark comparisons for the highest response to pedohebephilic stimuli

	df	λ	Non-central *t* (upper 5%/1%)	g_{cv} (α = .05)	g_{cv} (α = .01)	$g_{treatment}$	Sig.
Behavioural	182	4.223	5.954/6.696	0.441	0.496	0.798	**
with single case designs	207	4.503	6.230/6.969	0.433	0.484	0.866	**
Comprehensive	187	4.280	6.010/6.751	0.439	0.494	0.339	ns
Pharmacological	32	1.771	3.599/4.453	0.636	0.787	0.703	*
EMDR + Behavioural	11	1.038	3.049/4.160	0.919	1.254	0.749	ns

Table D-3

Benchmark comparisons for specific behavioural treatments

	df	λ	Non-central *t* (upper 5%/1%)	g_{cv} (α = .05)	g_{cv} (α = .01)	$g_{treatment}$	Sig.
Pedohebephilic Interests							
Olfactory aversion[a]	14	1.212	3.173/4.202	0.819	1.085	1.353	**
with single case designs	17	1.328	3.248/4.222	0.766	0.995	1.213	**
Covert & vicarious sensitization	74	2.711	4.477/5.258	0.517	0.607	0.649	**
Satiation	88	2.953	4.711/5.482	0.499	0.581	0.764	**
with single case designs	94	3.051	4.806/5.573	0.493	0.572	0.788	**
Positive conditioning + Extinction	39	1.980	3.791/4.624	0.599	0.731	0.604	*
Aversion + Extinction	97	3.098	4.852/5.619	0.490	0.568	0.663	**
Signalled punishment + Biofeedback	149	3.833	5.570/6.319	0.455	0.516	0.388	ns
Positive conditioning + Aversion + Extinction[b]	409	6.338	8.051/8.776	0.398	0.433	0.188	ns
Pedophilic Interest							
Satiation	32	1.798	3.628/4.485	0.632	0.781	1.081	**
with single case designs	35	1.878	3.699/4.544	0.617	0.757	1.116	**
Aversion + Extinction	57	2.384	4.166/4.965	0.547	0.652	0.296	ns
Teleiophilic Interest							
Olfactory aversion with single case designs	13	--	--	--	--	0.120	ns
Directed masturbation with single case designs	32	--	--	--	--	–0.304	ns
Satiation	16	--	--	--	--	–0.025	ns
with single case designs	21	--	--	--	--	–0.037	ns
Aversion + Extinction	57	0.936	2.643	0.350	--	0.225	ns
Positive conditioning + Aversion + Extinction[b]	431	2.574	4.238	0.204	--	0.195	ns

Table D-4

Benchmark comparisons for behavioural treatments in sexual offender against children subgroups

	df	λ	Non-central *t* (upper 5%/1%)	g_{cv} (α = .05)	g_{cv} (α = .01)	$g_{treatment}$	Sig.
Pedohebephilic Interests							
SOC-E	82	2.834	4.596/5.391	0.508	0.595	0.759	**
SOC-I	28	1.656	3.506/4.386	0.663	0.829	0.551	ns
SOC-FV	56	2.342	4.126/4.928	0.551	0.659	0.578	*
Adult SOC	320	5.599	7.317/8.046	0.409	0.450	0.705	**
Juvenile SOC	110	3.283	5.031/5.792	0.480	0.552	0.575	**
Teleiophilic Interest							
SOC-E	--	--	--	--	--	–0.484	ns
Adult SOC	--	--	--	--	--	–0.103	ns

Table D-5

Benchmark comparisons for taxon membership analyses

	df	λ	Non-central *t* (upper 5%/1%)	g_{cv} (α = .05)	g_{cv} (α = .01)	$g_{treatment}$	Sig.
Pedohebephilic Interests							
All treatments combined							
Taxon 1	326	--	--	--	--	0.121	ns
Taxon 2	150	3.846	5.583/6.331	0.454	0.515	0.436	ns
Taxon 3	226	4.716	6.441/7.178	0.428	0.476	0.741	**
Behavioural + Comprehensive							
Taxon 1	326	--	--	--	--	0.121	ns
Taxon 2	140	3.717	5.456/6.207	0.459	0.523	0.416	ns
Taxon 3	194	4.371	6.100/6.841	0.437	0.490	0.763	**
Behavioural treatments							
Taxon 1	15	--	--	--	--	0.305	ns
Taxon 3	170	4.093	5.826/6.570	0.446	0.502	0.684	**
Pedophilic Interest							
All treatments combined							
Taxon 1	326	--	--	--	--	0.091	ns
Taxon 2	40	--	--	--	--	0.297	ns
Taxon 3	61	2.465	4.243/5.036	0.539	0.640	0.952	**
Behavioural + Comprehensive							
Taxon 1	326	--	--	--	--	0.091	ns
Taxon 2	40	--	--	--	--	0.297	ns
Taxon 3	44	2.100	3.901/4.722	0.582	0.704	1.181	**
Behavioural treatments							
Taxon 1	15	--	--	--	--	0.293	ns
Taxon 3	44	2.100	3.901/4.722	0.582	0.704	1.181	**
Teleiophilic Interest							
All treatments combined							
Taxon 1	326	2.242	3.909/4.609	0.216	0.255	0.211	ns
Taxon 2	162	1.538	3.259/3.971	0.255	0.311	0.168	ns

Taxon 3	--	--	--	--	--	–0.196	ns
Behavioural + Comprehensive							
Taxon 1	326	2.242	3.909/4.609	0.216	0.255	0.211	ns
Taxon 2	162	1.538	3.259/3.971	0.255	0.311	0.168	ns
Taxon 3	--	--	--	--	--	–0.192	ns
Behavioural treatments							
Taxon 1	15	0.480	2.309/3.233	0.596	0.834	0.278	ns
Taxon 3	--	--	--	--	--	–0.192	ns