9-2019

Examining the Use of Hebephilia and Paraphilia Non-consent in Sexually Violent Predator (SVP) Evaluations

Nicole Graham
The Graduate Center, City University of New York

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

EXAMINING THE USE OF HEBEPHILIA AND PARAPHILIA NON-CONSENT IN

SEXUALLY VIOLENT PREDATOR (SVP) EVALUATIONS

By

Nicole Graham

A dissertation submitted to the Graduate Faculty in Clinical Psychology at John Jay College of

Criminal Justice in partial fulfillment of the requirements for the degree of Doctor of Philosophy,

The City University of New York

2019

Examining the Use of Hebephilia and Paraphilia Non-consent in Sexually Violent Predator (SVP) Evaluations

by

Nicole Graham

This manuscript has been read and accepted for the Graduate Faculty in Clinical Psychology at John Jay College of Criminal Justice, in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

_____ _____

Date Cynthia Calkins, Ph.D.

Chair of the Examining Committee

_____ _____

Date Richard Bodnar, Ph.D.

Executive Officer

Supervisory Committee:

Elizabeth Jeglic, Ph.D.

Rebecca Weiss, Ph.D.

Marcus Boccaccini, Ph.D.

Sandy Lewis, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Examining the Use of Hebephilia and Paraphilia Non-consent in Sexually Violent Predator (SVP) Evaluations

by

Nicole Graham

Advisor: Cynthia Calkins, Ph.D.

Sexually violent predator (SVP) legislation requires, in part, that an individual has a mental abnormality which causes difficulty in controlling sexual behavior. Previous research has found paraphilia not otherwise specified (NOS) to be one of the most prevalent diagnoses proffered in SVP evaluations. Since these studies, however, the Diagnostic and Statistical Manual (DSM) has updated the paraphilia NOS diagnosis in two ways. First, this diagnosis has been divided into two new diagnoses—other specified paraphilic disorder (OSPD) and unspecified paraphilic disorder. Second, OSPD requires an added specifier to indicate the individual's source of sexual arousal. To date, no study has systematically explored how the revision to paraphilia NOS has affected diagnoses within SVP evaluations. The current study explored the prevalence and diagnostic reliability of paraphilic disorders and associated specifiers in a sample of 190 adult men evaluated for SVP civil commitment. Results indicated that OSPD was the second most common paraphilic disorder, next to pedophilia, proffered in these SVP evaluations. However, there was poor to fair agreement between evaluators in providing this diagnosis. Additionally, while 'non-consent' and 'hebephilia' were the most commonly used specifiers, there was little consistency in which specifiers were used; and evaluators appear to be using an idiosyncratic approach to determine which labels to apply to OSPD. Given that the presence of a mental abnormality is the cornerstone to the constitutionality of SVP commitment, diagnostic practices

should be based in reliable and valid techniques. Implications for research, practice, and

legislation are discussed.

ACKNOWLEDGEMENTS

Most of all, the path to my doctoral degree would not have been possible without the influence from my number one role-model, my mother. From an early age she instilled values of hard work, persistence, and independence. Her unconditional support for any decision I made—even those she knew would fail—taught me to aim high and never give-up in the face of adversity.

Finally, I'd like to thank the city of New York. Thank you to all the coffee shops where I extended my stay working on this dissertation; thank you for the diverse opportunities to build my clinical repertoire; thank you for all the stalled trains; the untimely "sick-passenger;" the fifth-floor walk-ups; and endless rodent entertainment. The trials and tribulations of being a graduate student in the Big Apple are unique and overwhelming. But, you built me into a stronger (albeit, less patient) person. Thank you, NYC, for if you can make it here, you can make it anywhere.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

To date, Sexually Violent Predator (SVP) statutes have been implemented by twenty states as well as the federal government. These statutes allow for the indefinite, post-sentence civil commitment of individuals deemed to be high-risk sexual offenders. The purpose of SVP commitment is two-fold. SVP commitment is purposed to protect the community by detaining the highest-risk offenders who have committed sexually-based crimes, while also to provide mental health treatment aimed at rehabilitating offenders should they eventually return to the community (18 U.S.C. § 4248).

Sexually Violent Predator Commitment Criteria

Although the specific criteria for SVP civil commitment varies by state, in general, for an individual to be eligible for commitment he or she must have 1) a history of sexual offending, 2) a mental abnormality or personality disorder that 3) makes it difficult or impossible to control sexual behavior, and 4) be deemed as having a high likelihood to re-offend (*Kansas v. Crane*, 2002, *Kansas v. Hendricks*, 1997).

The first prong of these criteria is that an individual must have a history of at least one sexually violent offense. The term "sexually violent" is used by all jurisdictions except Minnesota (which uses the term, "engaged in a course of harmful sexual behavior") and North Dakota (which uses the term, "engaged in sexual predator behavior"). The definition of "sexually violent" varies by state but typically includes any type of sexual contact with a child, any type of coercive sexual contact with an adult, and sexual assault involving a weapon. In addition, "sexually motivated" non-sexual offenses, such as murder or kidnapping, can also meet this criterion (Doren, 2002). Determined through criminal record, there is little room for subjectivity with this requirement.

The second, third, and fourth requirements of the SVP criteria, however, are less objectively defined and are open for more debate. The second and third prongs are typically worded sequentially and require that the individual has a mental abnormality or personality disorder that predisposes him to some degree of volitional impairment that makes it difficult to control sexual behavior. States vary in the terminology, sometimes exchanging 'mental abnormality' for 'mental illness,' 'mental disorder,' or 'behavioral disorder' (DeMatteo, Murphy, Galloway, & Cox, 2015). Overall, however, most states' definition is modeled after *Hendrick*s (*Kansas v. Hendricks*, 1997), requiring that an individual suffer from "a congenital or acquired condition affecting the individual's emotional or volitional capacity which predisposes the person to commit sexually violent offenses to a degree constituting that such a person is a menace to the health or safety of others" (Kan. Stat. Ann. § 59-29a02; for a review of each state's definition, see DeMatteo et al., 2015). In sum, the second and third prongs include two requirements: (1) that the individual has a mental condition and (2) that this defect predisposes the individual to commit future criminal sexual acts (Miller, Amenta, & Conroy, 2005). No further legal guidelines exist as to what this mental condition should entail. Although legislation does not specify that a DSM-defined mental disorder need to satisfy this prong, research exploring the prevalence of diagnoses utilized in SVP evaluations (e.g., McLawsen, Scarlora, & Darrow, 2012) suggest evaluators employ the Diagnostic and Statistical Manual of Mental Disorders (DSM) to form their opinions. On the other hand, the DSM-5 cautions about the potential risks of mental diagnoses being "misused or misunderstood" (American Psychiatric Association [APA], 2013; p. 25) within forensic settings. While it is stated that "the use of an established system of diagnosis enhances the value and reliability of the [mental disorder] determination;" (APA, 2013; p. 25) it is also stated that a DSM-5 mental disorder is not synonymous with meeting requirements of a particular legal standard (e.g., SVP civil

commitment). Nevertheless, there is no guidance about what type of mental disorders or symptoms are permissible within this legislation. This debate is central to this manuscript and is addressed in further detail below.

The third criterion requires that the mental abnormality identified in prong two affects the emotional or volitional capacity of an individual which influences that person to commit future sexually violent behavior. However, the courts have not provided an operationalized definition of 'emotional or volitional capacity,' nor have they provided methods to assess such capacity. Thus, SVP evaluators must use their own discretion in defining and measuring these constructs. In one study, Mercado, Bornstein and Schopp (2006) surveyed legal professionals, psychologists, and mock jurors about how they assessed this capacity. Participants in this study reported that history of sexual violence, lack of offense planning, offender verbalization of control, and context of the SVP hearing contributed to their decision on emotional or volitional capacity. These findings are informative in understanding what SVP evaluators rely on in making determinations about volitional control, yet, there has been no empirical evidence suggesting these factors are associated with emotional or volitional capacity.

The fourth prong of SVP commitment requires that the individual is considered to have a high likelihood of sexual re-offense. One of the significant issues with this criterion is that the term "likely" has gone largely undefined. Knighton, Murrie, Boccaccini, and Turner (2014) reviewed case law of those jurisdictions that employ SVP statutes and found that some states (e.g., Florida; Iowa; Missouri; Nebraska; Washington; Wisconsin) identify "likely" quantitatively (i.e., risk of recidivism "must exceed 50%), while other states suggest a qualitative definition (i.e., risk of recidivism is "highly likely," or, will occur "more often than not"). Still, some jurisdictions have no case law providing any sort of definition.

SVP evaluators typically rely upon evidenced-based actuarial risk assessment measures in determining if an individual is likely to sexually recidivate, although this has not been specified by the legislation. These methods of assessment measure dynamic and static factors that have been found to be associated with sexual recidivism. Although the tools' respective manuals indicate high levels of inter-rater reliability (i.e., Psychopathy-Checklist Revised [PCL-R; Hare, 2003], Static-99 [Hanson & Thorton, 2000] , Minnesota Sex Offender Screening Tool-Revised [MnSOST-R; Epperson, Kaul, Hout, Hesselton & Alexander, 1998]), studies examining their field reliability have shown less promising results (e.g., Boccaccini, Murrie, Caperton, & Hawes, 2009; Boccaccini, Turner, & Murrie, 2008; Edens, Boccaccini, & Johnson, 2010; Miller, Kimonis, Otto, Kline, & Wasserman, 2012; Murrie, Boccaccini, Turner, Meeks, & Woods, & Tussey, 2009). For example, Miller and colleagues (2012) found only modest levels of inter-rater reliability (0.60 for the PCL-R; 0.74 for the MnSOST-R; and .78 for the Static-99) among evaluators in Florida. The proposed explanations for these low levels of field-reliability include poor evaluator training, subjective characteristics requiring measurement (i.e., affective traits), and adversarial allegiance (Boccaccini et al., 2008; Edens et al., 2010; Miller et al., 2012; Murrie, et al., 2009). These conclusions are concerning as findings from the assessment tools contribute to decision making regarding the indefinite civil commitment of an individual.

The Mental Abnormality Criterion is Fundamental

As previously stated, the second prong requires that the individual suffer from a mental abnormality or personality disorder that impairs volition and predisposes the person to commit a sexual offense. Although no legal guidelines exist as to what this mental condition should entail, such a defect is integral to SVP civil commitment. As explained, SVP legislation allows for the indefinite civil commitment of an offender after he or she has already served a prison sentence. However, such preventative detention would not be constitutional if the offender did not present

with a mental condition increasing the likelihood he or she would commit future sexual violence. As such, despite the ambiguity in legislative requirements, the presence of a mental abnormality is fundamental to the constitutionality of SVP civil commitment.

Diagnoses used in SVP evaluations.

The most commonly utilized diagnoses for SVP civil commitment are paraphilic disorders. Specifically, pedophilia and paraphilia not otherwise specified (NOS) are used most frequently (Becker, Stinson, Tromp, & Messer, 2003; Elwood, Doren, & Thorton, 2010; Jackson & Richards, 2007; Janus & Walbek, 2000; Levenson, 2004; Lu, Freeman, & Sandler, 2015; McLawsen et al., 2012). Whereas pedophilia has more specific diagnostic criteria, paraphilia NOS include vague criteria that can mistakenly be used as a catchall category for clinicians who deem there to be an abnormal sexual problem but for which the individual does not meet diagnostic criteria of a specified diagnosis. Much of the research that had evidenced the frequent use of these diagnoses was based on the fourth edition of the DSM which utilized the term paraphilia NOS. The fifth edition of the DSM, which was published in 2013 (DSM-5; American Psychological Association, 2013), no longer carries a paraphilia NOS category, rather, this classification has been modified into two distinctions, 'Other Specified Paraphilic Disorder (OSPD),' and 'Unspecified Paraphilic Disorder (UPD).' All three of these categories are largely similar in that they allow the clinician to provide a diagnosis for an abnormal sexual interest that causes marked impairment but does not meet any of the eight DSM-5 paraphilic diagnoses (e.g., pedophilia, sexual sadism, frotteurism). Where the fifth edition varies, however, is with the requirements for OSPD. For this diagnostic category, the clinician is expected to provide a specifier indicating the specific sexual interest that individual has that is abnormal and causes impairment or distress.

To date, nine studies have explored the prevalence of diagnoses among sexual offenders who were evaluated for civil commitment (Perillo, Spada, Calkins & Jeglic, 2013) and, or, ultimately committed (Becker et al., 2003; Elwood et al., 2010; Jackson & Richards, 2007; Janus & Walbek, 2000; Levenson, 2004; Lu et al., 2015; McLawsen et al., 2012). As noted, these studies were conducted while the fourth edition of the DSM was being utilized by clinicians and no research has explored the prevalence of diagnoses using the fifth edition of the DSM. As such, it is unknown how often OSPD and UPD are being used, or what specifiers are being used for OSPD. Interestingly, however, three of these previous studies (Elwood et al., 2010; Jackson & Richards, 2007; McLawsen et al., 2012) parsed out diagnoses of paraphilia NOS into 'paraphilia NOS, non-consent' and 'paraphilia NOS, excluding non-consent.' Likewise, a review of SVP cases between 2008 and 2011 revealed that paraphilia NOS, non-consent has been used with increased frequency (King, Wylie, Brank, & Heilbrun, 2014). More recently, several *Frye* hearings in the State of New York (e.g., *Matter of State of New York v. Ralph P.,* 2016; and *Matter of State of New York v. Jason C.,* 2016) have suggested the terms 'non-consent' and 'hebephilia' are being used as specifiers for the OSPD diagnosis.

What is Hebephilia and Paraphilia Non-Consent?

Hebephilia does not have a formal definition; however, it generally refers to a sexual preference for pubescent-aged adolescents. It is specifically distinct from pedophilia, which is the sexual preference for prepubescent-aged children, and from teleiophilia, the sexual preference for adults. Stephens, Seto, Goodwill, and Cantor (2016) note that hebephilia has been conflated with an interest in older adolescents (e.g., 15 – 17-year-olds), however, hebephilia specifically refers to the interest in youth who are in Tanner stages 2-3. The Tanner stages describe the primary and secondary physical features of sexual development from childhood to adulthood (e.g., size of breasts or testes, development of pubic hair; Tanner, 1990). Those in the

6

2nd and 3rd stages are beginning to show *some* secondary sexual characteristics which would indicate the initial growth of pubic hair as well as budding breasts; versus older adolescents whose sexual development more closely resembles that of an adult. Typically, these stages refer to those who are 11 – 14 years old, however, age is not a definitive factor as sexual development varies among individuals.

Paraphilia non-consent is a construct most in line with the crime of rape. Like hebephilia, paraphilia non-consent has no standard definition. In general, it typically refers to sexual arousal to coercive, sexual contact with non-consenting individuals (e.g., Wakefield, 2011). This construct has been referred to by several different terms. Paraphilia non-consent appears to be the most used term, however, paraphilic coercive disorder and biastophilia have also been used. By some, these terms appear to be referring to the same overall construct and have even been used interchangeably in the same paper (e.g., Knight, 2010). Others (e.g., Money, 1999; Thorton, 2010) have defined differences in the two. Paraphilic coercive disorder has been explained as a sexual arousal to the coercive nature of the rape. Whereas biastophilia has been described as a sexual arousal to the coercive nature *and* to the victim's terror and resistance. For the purpose of this study, however, paraphilia non-consent will be the only term utilized but represents more broadly the sexual arousal to coercive sexual interactions.

Both of these constructs—hebephilia and paraphilia non-consent—have stirred much debate[1]. Over the past decade there has been growing interest, but also concern, over use of the labels hebephilia and paraphilia non-consent in the courtroom. Although hebephilia had its origin dating back to the 1950's, the resurgence of interest in this label has taken place during the last

---

[1] The following works present opposing sides to the debate of hebephilia and paraphilia non-consent as mental health diagnoses: Blanchard (2009; 2010); DeClue (2006; 2009); Frances & First (20011a; 2011b); Franklin (2009; 2010); Green (2010); Knight (2010); Moser (2009); Plaud (2009); Prentky, Coward & Gabriel (2009); Quinsey (2010); Rind (2013); Rind & Yuill (2012); Stern (2010); Thorton (2010); Tromovitch (2009); Tucker & Brakel (2012); Wakefield (2011); Zander (2007; 2008; 2009).

fifteen or so years and mostly within the realm of SVP commitment (Franklin, 2010). Similarly, although there have been decades of research attempting to understand different sources of sexual arousal for individuals who commit rape, efforts to pathologize rape (e.g., pedophilia non-consent) has been used almost exclusively in the SVP context (King et al., 2014). Thus, it has been argued that the empirical interest does not seem to be guided by an overall curiosity in paraphilic diagnoses, but rather to establish these constructs as mental disorders for the purpose of civil commitment of sexually dangerous persons (e.g., Frances & First, 2011b; Franklin, 2010; King et al., 2014; Moser, 2009; Tromovitch, 2009, Zander, 2008, 2009).

The Use of Hebephilia and Paraphilia Non-Consent in SVP Evaluations

The constructs of hebephilia and paraphilia non-consent were rejected from the DSM-5 as a mental disorder construct, rejected as specifiers under the OSPD diagnosis, and also rejected in the appendix as an area warranting further study (Frances & First, 2011b; King et al., 2014; Zander, 2010). The rejection of these constructs as mental disorders was based on arguments that the empirical research which has attempted to establish these constructs as distinct mental health diagnoses is minimal and the methodology utilized in these small number of studies have been fraught with issues. Further, it was argued that findings supporting the reliability and validity of these constructs is weak and any evidence relative to the etiology and prevalence of these constructs are in the nascent stages. Interestingly, recent Frye[2] hearings have ruled that a diagnosis of hebephilia (*Matter of State of New York v. Ralph P.,* 2016) and paraphilia non-consent (*Matter of State of New York v. Jason C.,* 2016; *Matter of State of New York v. Kareem M.,* 2016) was inadmissible for SVP commitment *not* because it was not incorporated into the DSM-5 (as the petitioner also argued in *Matter of State of New York v. Jason C.,* 2016), but

---

[2] A Frye hearing is used to determine whether evidence that was obtained is generally accepted by experts in that field for which it is being used (*Frye* v. *United States,* 1923). This standard has been superseded by *Daubert* v. *Merrell Dow Pharmaceuticals* (1993) in the majority of states, though is still relevant in New York.

because there was not a general consensus among those in the field of the acceptability of this construct as a diagnosis.

Although the court in these specific cases ruled that hebephilia and paraphilia non-consent were not generally accepted diagnoses within the field of psychology, other courts have accepted the admissibility of such diagnoses (e.g., *The People of the State of Illinois, v. Kevin Stanbridge,* 2012*)*. Further, pilot data has suggested that paraphilia non-consent was the most commonly utilized specifier within SVP civil commitment evaluations—with hebephilia being the second most prevalent (Graham & Calkins, 2017). Thus, there may be some disagreement in the field about whether these constructs are 'generally accepted.' However, current literature does not address this possibility. As such, the first aim of this study is to examine the prevalence in which hebephilia and paraphilia non-consent are being used as specifiers for OSPD in SVP evaluations.

Inter-Rater Reliability of Hebephilia and Paraphilia Non-Consent

SVP evaluations have high-stake outcomes given that they can impose on the civil liberties of individuals. Therefore, it is imperative that these evaluations are conducted in the utmost reliable and valid manner feasible. Previous research has suggested there may be questionable reliability of the diagnoses provided in SVP evaluations. For example, Perillo and colleagues (2014) examined the inter-rater reliability of diagnoses provided by independent evaluators in 375 SVP cases. Their results suggested poor (.23) to moderate (.55) agreement across the diagnoses given by clinicians in this context. Further, evaluators' ability to reliably diagnose paraphilia NOS was poor (*kappa* = .33; see also, Packard & Levenson, 2006). Research in this area, however, has not been conducted with the newest edition of the DSM. Thus, it is unknown if the new paraphilia OSPD criteria—which aimed to improve diagnostic clarity—has improved inter-rater reliability. Further unknown is whether SVP evaluators reliably use the

same specifier. Thus, the second aim of this study is to explore the inter-rater reliability of paraphilia OSPD generally, and hebephilia and paraphilia non-consent, specifically.

Are There Differences Among Evaluators?

It is also unknown *who* is using these specifiers. Clearly there is a debate in the empirical literature as to whether such constructs exist as a mental disorder, as was also evident in recent Frye hearings. Relevant to this question is the work conducted by those who have explored field-reliability of risk assessment measures in SVP evaluations (see above; Boccaccini et al., 2009; Boccaccini et al., 2008; Edens et al., 2010; Miller et al., 2012; Murrie et al., 2009). The findings from these studies have evidenced poor reliability in risk assessment scores within SVP contexts; and that this poor reliability may be due not only to adversarial allegiance (i.e., prosecution-retained evaluators deeming individuals to be higher risk than those retained by the defense) but simply due to differences in evaluators. For example, in a sample of 321 SVP evaluations assessed by 1 or more of 20 state-contracted evaluators, over 30% of the variability in PCL-R scores were attributed to differences among evaluators. That is, some evaluators tended to consistently rate sex offenders higher or lower risk than other evaluators did. These findings suggest there are idiosyncratic differences amongst evaluators—perhaps due to training, experience, or even personality (see, Miller, Ruffino, Boccaccini, Jackson, & Murrie, 2011)—that influence an evaluator in his or her approach to scoring risk assessment tools. Thus, it is worth questioning whether there are some evaluators who tend to use the constructs of hebephilia and, or, paraphilia non-consent, or whether utilization of these specifiers is employed more generally across all SVP evaluators. If the latter is true, this may have important implications for policy regarding the 'general acceptance' prong in admissibility determinations.

Current Study

      SVP evaluations have vital outcomes. Their consequences can impose on the civil liberties of the individual being evaluated; effect the safety of our communities; and demand increased expenditures within our country. Therefore, it is in the interest of all that these evaluations are conducted with the utmost integrity. However, a review of the literature has identified some areas of concern in the practice of SVP evaluations. Therefore, this study seeks to address four aims to help address these limitations.

      Aim 1. As discussed, SVP statutes require that, for civil commitment, the sexual offender must have a mental abnormality that makes it difficult or impossible to control sexual behavior. However, there has been little legal guidance as to what this mental abnormality should entail. Previous research has suggested that paraphilia NOS is one of the most frequently-used categories in SVP evaluations (e.g., Elwood et al., 2010; Perillo et al., 2014). However, since these studies have been published, the DSM-5 modified the paraphilia NOS categories into two distinct categories: other specified paraphilic disorder (OSPD) and unspecified paraphilic disorder (UPD). The research in the field has not been updated with these modified diagnostic categories. Thus, the first aim of this study is to explore the frequency in which OSPD and UPD are used in SVP evaluations.

      Hypothesis 1. Based on initial pilot data and previous research, it is hypothesized that OSPD will be one of the most frequently used diagnostic categories used in SVP evaluations. It is further hypothesized that UPD will be used with regularity, but not as frequently as OSPD.

      Aim 2. With the newly created OSPD diagnostic category, the DSM-5 requires clinicians to identify the source of problematic sexual arousal for the evaluee. There is no empirical evidence informing the field what specifiers are used here. Earlier research has suggested SVP evaluators utilize paraphilia non-consent as a specifier to relate to a sexual attraction to non-

consensual sex. Anecdotal evidence from Frye hearings and preliminary research has suggested, in addition to paraphilia non-consent, that SVP evaluators also utilize the label hebephilia, referring to a sexual attraction to pubescent youth. Use of these constructs, however, has stirred much debate and recent court rulings have deemed hebephilia and paraphilia non-consent inadmissible due to a lack of general acceptance within the field. Yet, pilot data suggest that hebephilia and paraphilia non-consent may be used with more frequency than thought (Graham & Calkins, 2017). Thus, the second aim of this study is to explore the extent to which hebephilia and paraphilia non-consent are used as specifiers for OSPD within SVP evaluations.

Hypothesis 2. Based on the initial research that was conducted, it is hypothesized that hebephilia and paraphilia non-consent will be the most frequently utilized specifiers for the OSPD category.

Aim 3. Previous research has evidenced poor reliability of diagnoses utilized in SVP evaluations; this has been specifically apparent for the paraphilic diagnoses (Packard & Levenson, 2006; Perillo et al, 2014). Thus, the third aim of this study is threefold. For one, it is to update previous research by exploring if the reliability of paraphilic diagnoses has improved. Second is to understand whether the newly adapted OSPD category demonstrates better inter-rater reliability compared to paraphilia NOS. Additionally, no study has explored the ability of evaluators to consistently utilize the same specifier within the OSPD category (e.g., hebephilia, paraphilia non-consent), which this study also aims to do.

Hypothesis 3. Given previous research suggesting poor inter-rater reliability estimates of paraphilia NOS, and more recent initial findings (e.g., Graham & Calkins, 2017), it is hypothesized that the inter-rater reliability for any paraphilic diagnosis, as well the inter-rater reliability for OSPD and UPD generally, and the OSPD specifiers more specifically, will also be poor.

Aim 4. Research within this area has suggested that evaluators may apply idiosyncratic approaches to SVP evaluations. As previously mentioned, Boccaccini and colleagues (2008) demonstrated that some evaluators consistently assign higher or lower PCL-R scores than others, even when the examinee presents no notable differences. Similarly, Miller and colleagues (2011) showed that evaluators who described themselves as being more agreeable on a personality measure rated examinees as less psychopathic on the interpersonal factor of the PCL-R. More relatedly, practitioner differences are clearly evident given opposing sides in recent Frye hearings. Thus, the fourth aim of this study is to understand if there are a subset of evaluators who employ OSPD diagnoses more frequently than other evaluators. Further, this study will seek to understand if there are a subset of evaluators who use the hebephilia and paraphilia non-consent specifiers.

Hypothesis 4. Despite the abovementioned evaluator differences, it is hypothesized that OSPD and UPD in general, and more specifically, hebephilia and paraphilia non-consent, are being used across the board and there will be no subset of evaluators that provide these diagnoses.

CHAPTER 2: METHOD

Data Collection

Data for the present study were provided by the Florida Department of Children and Families, Sexually Violent Predator Program (herein, SVPP). The SVPP houses all records of offenders who were evaluated under Florida's SVPP. The primary investigator and three research assistants coded the files in two data collection trips (February, 2017 and December, 2017).

The primary investigator was the author of this manuscript; a doctoral student completing a Ph.D. in clinical psychology with an emphasis in forensics. The primary investigator began conducting SVP research during her undergraduate career and was a member of the Sex Offender Research Lab (SORL) at John Jay College. The research assistant who helped during the first data collection trip was completing a master's degree in forensic psychology and also a member of SORL. The research assistants available for the second data collection trip were both first year doctoral students. One was enrolled in the same clinical psychology doctoral program as the primary investigator; and the other was enrolled in a different clinical psychology Ph.D. program, but whom already completed a master's degree in forensic psychology and was a prior member of SORL.

Florida's Sexually Violent Predator Program.

Offenders who have a "sexually violent offense" are referred by the Department of Corrections, Department of Juvenile Justice, and the Department of Children and Families to the SVPP multidisciplinary team (MDT) 545 days before their release (or as soon as possible if incarceration period is shorter). All referrals are screened for risk by the MDT and they may then refer the offender for a face-to-face evaluation. These contracted evaluations are usually performed by doctoral level psychologists who are required to opine whether the offender meets

14

the state's definition of a sexually violent predator. Specifically, by Florida definition, a sexually violent predator is someone who "has been convicted of a sexually violent offense; and suffers from a mental abnormality or personality disorder that makes the person likely to engage in acts of sexual violence if not confined in a secure facility for long-term control, care, and treatment" (Florida § 394.912). The MDT initially requests one evaluation. If that evaluation indicates that the offender does meet commitment criteria then a request for a second independent evaluation is made. Occasionally the MDT will request a second evaluation even if the first evaluation results in the opinion that the offender does not meet criteria. This may occur when the MDT concludes the first evaluation did not sufficiently answer an important question or perhaps when new information surfaces. The MDT reviews these evaluations and makes a final recommendation. If a recommendation for civil commitment is made, then this opinion is sent to the state attorney and a petition may be filed.

     Sample.

     Given the aims of this study, there were several inclusion criteria. First, given that one aim of this study is to explore the prevalence of the paraphilic categories adopted by the DSM-5, only those evaluations conducted after this edition was published (May 2013) were included. That is, this sample only included evaluations conducted between May 2013 and June 2017[3] ($n =$ 611). Additionally, this sample included only those files in which two evaluations were conducted ($n = 255$). As mentioned, there are instances in which the individual only receives one evaluation; however, given that one of the aims of this study was to explore the reliability in mental health diagnoses among evaluators, it is necessary the records included only those in which two evaluations were conducted. To put into context, those files that meet this study

---

[3] This month was determined based on the time point in which data for Figure 1 was obtained.

criteria are approximately 42% ($n = 255$) of those individuals who were referred by the MDT for an SVP evaluation (see Figure 1).

The initial sample was to include 255 male sexual offenders who had been convicted of at least one felony and were evaluated for civil commitment under Florida's SVPP between the years of 2013 and 2017. However, the final dataset included a total of 190 cases, or 380 evaluations. The final number of cases was lower than anticipated as 65 cases (25%) were excluded from the final sample because at least one evaluator did not use the DSM-5. As stated, the current study only considered assessments that utilized the fifth edition due to its modification of the paraphilia NOS category.

The entire sample of sexual offenders were males above the age of 18 ($n = 190$). Offenders were identified as White (53%), Black (42%), Hispanic (4%), or other ethnic minorities (1%). Offenders were incarcerated for a variety of sexual offenses and have a history of a sexually violent offense per Florida statute (e.g., sexual battery; lewd or lascivious act with or in presence of child; kidnapping or false imprisonment of a child involving sexual battery or lewd or lascivious acts; murder while engaged in sexual battery; Florida § 394.912). Further, 84% ($n = 160$) of the sample had a reported history of prior sexual offenses. The majority of offenders in this sample (70%) were referred for civil commitment by the MDT.

Evaluators.

The SVP evaluators are psychologists who, like the MDT members, have extensive expertise in this area. The evaluators, along with members of the MDT, are required to meet for a yearly conference where new research is presented and other issues that affect the evaluation process are discussed. These evaluators are contracted by the state and are assigned cases based largely on geographical proximity. Notably, because these evaluators are contracted by the state, and not by either side of the judicial process (e.g., defense or prosecution), any influence of

adversarial allegiance is *theoretically* non-existent; evaluators for each case are typically

determined based on geographic proximity and availability. In addition to this conference,

contracted evaluators are required to obtain continuing education training that is relevant to the

assessment of sex offenders

Of the 380 evaluations that were conducted, there were 21 distinct evaluators. All of the

evaluators were licensed psychologists in the state of Florida; most (67%) held a Ph.D., while a

third held a Psy.D. Thirteen of the evaluators were male (62%) and eight were female (38%).

Within this dataset, evaluators conducted a range of evaluations (Range 3- 39, M = 18).

Measures.

Data were extracted from each SVPP evaluation based on an established coding manual.

The primary investigator created this coding manual which was informed by the aims of this

study, previous research (e.g., Perillo et al., 2014), and the SVPP. The coding manual included

the variables described below and its operationalized definition. As most of the variables were

categorical, the coding manual identified set response options for each variable. The coding

manual was approved by the Institutional Review Boards of the City University of New York

and Florida Department of Children and Families.

The primary investigator reviewed the coding manual with each research assistant prior

to the data collection trip. Data were collected using an established spreadsheet which matched

the coding manual. For all categorical variables, drop-down options were provided in the

spreadsheet for coders to select the appropriate response; this helped to increase consistency in

data collection. Twenty cases (10%) were double-coded to assess for inter-rater reliability;

results for each variable are discussed below.

Offender demographics. The offender's race/ethnicity and year of birth were obtained.

There was 100% agreement between coders in obtaining this information.

Index offense. The index offense is the offense under which the offender is incarcerated. The year of this crime, the nature of this offense (i.e., molestation, adult sexual assault), and the victim(s)' age(s) were coded. Victim's age was coded categorically as a) 10 and under (see footnote regarding victims of 11 years old)[4]; b) 12; c) 13 – 14; d) 15 – 16; e) 17 and older; or f) multiple ages—which suggested the index offense included victims across multiple age groups.

Index victim age. The victim's age at the time of the index offense was coded as identified above (see also, footnote 2). There was 85% agreement ($n = 17$) between coders in obtaining this information.

Prior sexual offenses. Any prior sexual offenses were also coded. The information gathered for the prior sexual offenses was the same as that gathered for the index offense. There was 80% agreement ($n = 16$) between coders in obtaining prior victims' ages. However, given additional qualitative information that was collected, discrepancy between two of these cases was able to be remedied.

Preferred victim age. The offender's preferred victim age was calculated after data collection. If the age of the victim in the offender's index offense was within the same category as the victim(s)' age(s) in the prior sexual offenses, his preferred victim age fell within that category. If the ages varied across index and priors, the offender's preferred victim age was coded as 'multiple,' indicating he offended against victims in multiple age groups. If the offender had no prior sexual offenses, his preferred victim age group matched the age of the victim in his index offense.

---

[4] Given a difference in the way data was coded between two data collection points, victims who were 11 years old may have been excluded in some analyses. During the first data collection, victim's age was initially coded as 12 and under; during the second data collection victim's age was coded as a) 10 and under; b) 11; c) 12; d) 13- 14; e) 15-16; f) 17 and older; or g) multiple. When data from the first collection were recoded to fit the updated coding scheme, 43 cases could not identify whether the index victim age was 10 and under or 11 or 12. Similarly, for prior offense victim age, 17 cases could not be identified.
Of note, in the second data collection, neither index nor prior offenses included victims 11 years old.

SVPP Evaluations. Each SVPP evaluation was coded for the evaluator name, his or her gender and educational degree. Evaluator names were later assigned a non-identifiable specifier (e.g., letters A-U). Evaluations were specifically coded for the date of evaluation, whether a face-to-face evaluation was conducted—as some offenders reject the evaluation, what diagnoses were provided, whether the evaluator recommended the offender for civil commitment, and risk assessment scores. Notably, if a diagnosis of OSPD was provided, the specifier linked to this diagnosis was coded. In some instances the evaluator provided a specifier for UPD; this was also recorded. There was 100% agreement between coders in collecting this information.

CHAPTER 3: RESULTS

Aim 1. Prevalence of Diagnostic Categories

The most common diagnoses found in this sample were the paraphilic disorders, followed by personality disorders; substance-use disorders were also diagnosed in high frequency. Evaluators provided a paraphilic diagnosis for the majority ($n = 296$; 78%) of offenders they evaluated (see Table 1). Pedophilia was the most common paraphilic diagnosis ($n = 114$; 30%), followed closely by OSPD ($n = 107$; 28%). UPD was diagnosed in higher frequency ($n = 66$; 17%) than the remaining paraphilic diagnoses (e.g., exhibitionism, sexual sadism, and voyeurism).

Aim 2. Specifiers of OSPD

As noted, OSPD was the second most prevalent paraphilic diagnosis. Out of the 380 evaluations conducted, this diagnosis was offered 107 times (28%). Per the DSM-5, clinicians are to provide a specifier to indicate the source of sexual arousal. Of the 107 times this diagnosis was offered, clinicians provided a specifier 87% of the time ($n = 93$). While UPD does not require a specifier, clinicians provided one 18% of the time ($n = 10$). The specifiers offered for OSPD were far-ranging (see Table 2). The distinct label of 'non-consent' was provided 42 (39%) times. Similar labels such as "non-consenting sex;" and "non-consenting persons" were also used. The distinct label 'hebephilia' was provided seven times (7%). However, many of the specifiers provided could be presumed to mean 'non-consent' or 'hebephilia.' For example, on five occasions evaluators used the term "biastophilia" to specify OSPD which is often used synonymously for 'non-consent.' In an effort to fully understand these diagnostic practices, four variables were created (see Table 2).

A variable for 'non-consent' was created based on cases in which the evaluator provided a label utilizing the word 'non-consent' ($n = 61$). A 'non-consent combined' variable was created

to include all the 'non-consent' specifiers, as well as labels presumed to depict 'non-consent' (e.g., biastophilia; paraphilic rape; $n = 68$). Therefore, there were seven instances in which a label without the word 'non-consent' was added to this combined category (see Table 2). Similarly, a 'hebephilia' variable included only those cases in which the evaluator used the distinct language of 'hebephilia' ($n = 7$). The 'hebephilia combined' included nine more labels presumed to address hebephilic preferences (e.g., 'sexual activity with an adolescent;' see Table 2). Table 3 reports the prevalence of these variables.

To note, there were nine instances a label was used which did not appear to align with 'non-consent' or 'hebephilia' (e.g., 'bestiality;' 'sexting'); these labels were not included in subsequent analyses but can be reviewed in Table 2 and Appendix A.

Aim 3. Reliability of Diagnostic Categories

The consistency among evaluators' diagnostic tendencies was explored using several statistics (see Packard & Levenson, 2006) and reported in Table 4. First, kappa coefficients were calculated to measure inter-rater reliability for each diagnostic category and specifier. The Bloom, Fischer, and Orme (1999) standard for kappa agreement was used ("poor" = below 0.60, "fair" = 0.60 – 0.74, and "good" = 0.75 and above) to assess for poor to good reliability. With the exception of pedophilia (*kappa* $= 0.78$, $p < 0.01$), all paraphilic diagnostic categories exhibited poor reliability (*kappas* ranged $0.16 – 0.58$, $p < 0.05$). Kappa is advantageous in that it provides a measure of agreement that is beyond what would be expected by chance alone (Sim & Knight, 2005). That said, kappa coefficients can be distorted as this statistic is influenced by low base rates of a disorder, as well as disproportionate levels of agreement between evaluators. Given that prevalence of paraphilic disorders is largely unknown; and that evaluators were more likely to agree on the *absence,* rather than presence of a disorder, other agreement statistics should also be considered.

Positive predictive values (PPV) were used to demonstrate the probability that both evaluators agreed on the presence of a given diagnosis or specifier, given that the first evaluator provided that diagnosis. Of the paraphilic disorders, PPV was strongest for pedophilia, with a 91% agreement when the first evaluator provided a diagnosis of pedophilia.

Negative predictive values (NPV) indicate the probability that both evaluators agree the diagnosis is *not* present, given that the first evaluator did not provide said diagnosis. NPV trends were consistent across all the paraphilic categories; when the first evaluator did not provide a specific paraphilic diagnosis, the second evaluator was also unlikely to diagnosis this specific disorder (NPVs ranged 0.75 – 0.99). PPV and NPV values, however, are also sensitive to base rates of a disorder—such that diagnoses with a low prevalence will have a lower PPV and a higher NPV (Riddle & Stratford, 1999).

Proportion of agreement between evaluators was also explored. Proportions of agreement are descriptive statistics that compute the percent of times the evaluators agreed overall (i.e., both evaluators agreed in diagnosing or not diagnosing a specific disorder); or agreed on the presence (positive proportion of agreement; PA+) or absence (negative proportion agreement; PA-) of a disorder. Overall, evaluators were likely to agree that a paraphilic disorder, of some type, was present (PA+ = 0.82). However, agreement on specific disorders ranged from 0 (i.e., no agreement at all) to 0.85. Consistent with the aforementioned findings, evaluators were most likely to agree on the presence of pedophilia (PA+ = 0.85); there was no agreement in a diagnosis of sexual sadism. Evaluators were more consistent in opining when a specific paraphilic disorder was *not* present (this proportion ranged from 0.78 – 0.99).

Other specified paraphilic disorder.

Evaluators agreed on the presence of OSPD 43% of the time ($n = 46$). Based on the kappa coefficient, evaluators demonstrated poor agreement on an OSPD diagnosis (*kappa* = 0.21, $p < 0.01$). There was less than chance agreement if the first evaluator rendered this diagnosis that the second evaluator would do the same (PPV= 0.48). On the other hand, if evaluator one *did not* provide this diagnosis, there was about 75% chance the second evaluator would not.

Paraphilia not otherwise specified (NOS).

To compare the results of this study to previous studies (see Table 4), a paraphilia NOS variable was computed by combining OSPD and UPD diagnoses. Compared to OSPD, this computed paraphilia NOS variable demonstrated improved diagnostic consistency between evaluators. While kappa was still considered poor (*kappa* = 0.27, $p < .001$), evaluators were much more likely to agree on the presence of this diagnosis (PA+ = 64%) compared to OSPD (PA+ = 43%) or UPD (PA+ = 30%) alone.

Non-consent.

Consistency in providing a 'non-consent' label was analyzed using both of the 'non-consent' variables as identified above. Overall, evaluators showed poor agreement in using a non-consent specifier, no matter how non-consent was defined.

Poor agreement in the use of the 'non-consent' (only) specifier (*kappa* = 0.17, $p < 0.05$) was demonstrated. There was 30% agreement between evaluators in providing this label. There was a 36% chance if evaluator one provided the 'non-consent' specifier that the second evaluator would do the same. On the other hand, there was an 84% chance if the first evaluator *did not* provide this specifier, the second evaluator also would not provide this specifier.

The 'non-consent combined' variable—which provided more inclusivity of labels—only slightly improved reliability between evaluators (*kappa* = 0.22, p < 0.05); proportion of agreement increased to 35% (versus 30%; see Table 4).

Hebephilia.

Evaluators also demonstrated poor agreement in the use of the 'hebephilia' specifiers (e.g., 'hebephilia only' and 'hebephilia combined'). Poor agreement in the use of the 'hebephilia' (only) label was demonstrated (*kappa* = 0.27, *p* < 0.01). There was 29% agreement between evaluators in providing this label. There was a 20% chance if evaluator one provided the 'hebephilia' specifier that the second evaluator would do the same. On the other hand, there was a 99% chance if the first evaluator *did not* provide this specifier, the second evaluator also would not provide this specifier.

The 'hebephilia combined' variable did not improve reliability between evaluators (*kappa* = 0.24, p < 0.001); proportion of agreement decreased to 25% (versus 29%; see Table 4).

Qualitative exploration.

To understand how discrepant evaluators were in their diagnostic tendencies, each case was qualitatively explored (see Appendix A). Evaluators did not agree on an OSPD diagnosis 61 times (57%). That is, evaluator one provided an OSPD diagnosis 36 times when evaluator two did not; and evaluator two provided an OSPD diagnosis 25 times when evaluator one did not. When the two evaluators did not agree, the opposing evaluator provided the UPD diagnosis 31% of the time (*n* = 19). Of the remaining 42 cases, a paraphilic disorder, of some type, was provided 18 times (43%). These other paraphilic disorders were typically pedophilia (*n* = 12; 67%) or sexual sadism (*n* = 5; 28%).

Commitment recommendation.

As described above, each state-contracted SVP evaluator must make a recommendation about whether the offender should be civilly committed. If the evaluator opines the offender should be committed a second evaluation occurs. Evaluators agreed more often than not on their commitment recommendation (PA = 63%; $n = 119$). Evaluators were more likely to agree to commit the offender (PA+ = 74%) rather than disagree (PA- = 37%).

The psychologist performing the second evaluation is not formally informed a first evaluation has been conducted, nor are they privy to the first evaluation report. However, it is possible the second psychologist becomes anecdotally aware (e.g., the offender reports it) a first evaluation occurred. Knowing a first evaluation occurred—which likely resulted in a recommendation to commit—may influence the diagnostic tendencies of the second evaluator. As such, additional reliability analyses were conducted to explore this potential influence (see Table 5).

Reliability of evaluator's OSPD diagnostic tendencies were explored when the evaluators 1) agreed on their commitment decision ($n = 120$, 63%); 2) did not agree on their commitment decision ($n = 69$, 36%); and 3) both recommended the offender for civil commitment ($n = 100$, 52%). Overall, consistency between evaluators improved when they both agreed on the commitment decision and when they both recommended civil commitment. For example, in the full study, proportion of agreement (PA+) for a diagnosis of OSPD was 43% (*kappa* = .21, *p* < .01). However, when both evaluators agreed on their commitment decision the proportion of agreement increased to 52% (*kappa* = .29, *p* < .01); and when they both recommended to civilly commit the offender, proportion of agreement increased to 54% (*kappa* = .28, *p* < .01). On the other hand, when the evaluators disagreed on commitment decisions, reliability for OSPD decreased (PA+ = 20%; *kappa* = .04, *p* = .08). Findings for 'non-consent' and 'hebephilia' were

similar—evaluators were more consistent in providing these labels when they agreed on commitment recommendations; these results are depicted in Table 5.

Aim 4. Who uses these specifiers?

To try to understand who uses the OSPD diagnosis in general, and the specifiers of hebephilia and paraphilia non-consent, several analyses were conducted. As noted, there were 190 distinct SVP cases. For each of these cases, two evaluations were conducted. As such, analyses were conducted both by total number of evaluations ($N = 380$), as well as total distinct SVP cases ($N = 190$).

Overall, the diagnosis of OSPD was proffered in 28% of the evaluations ($n = 107$). However, evaluators' use of this disorder ranged from 0% to 67% (see Table 6). Notably, out of the 21 evaluators, 20 evaluators used the OSPD diagnosis at least once. One evaluator never used this diagnosis.

Non-consent.

In looking at all the evaluations ($N = 380$), the use of the specifier 'non-consent' was varied. To be inclusive, the 'non-consent combined' variable was explored in all aforementioned analyses. As noted, 'non-consent' terminology was used 68 times out of the 107 times OSPD was diagnosed (64%, see Table 3). It appears, on average, this specifier was used in half of the evaluations conducted (52%); however, this statistic is somewhat misleading. Rather, there were 13 evaluators who provided the non-consent specifier in 50% or more of their evaluations. Whereas five evaluators never used this descriptor (see Table 6).

In an attempt to elucidate differences amongst evaluators who use these specifiers, several considerations were made. First, evaluators were categorized based on the frequency of times he or she used 'non-consent.' Three categories were created which included 'never' used this specifier ($n = 5$), 'sometimes' used this specifier (e.g., used non-consent as a specifier in 1-

61% of the OSPD diagnoses proffered; $n = 5$) and 'often' (e.g., used non-consent as a specifier in 62 - 100% of OSPD diagnoses; $n = 10$).

Little information – apart from evaluator degree, gender, and number of evaluations completed – was available to differentiate evaluators. As noted, 13 psychologists held a Ph.D., and eight held a Psy.D. Evaluator degree did not differentiate whether one was more or less likely to provide a 'non-consent' specifier $X^2 (2, N = 20) = 2.42, p = 0.30$. Similarly, evaluator gender did not influence whether one provided a 'non-consent' specifier $X^2 (2, N = 20) = 0.51, p = 0.77$. Finally, the proportion of evaluations each clinician conducted was explored. As noted, evaluators in this sample conducted a range of $3 – 39$ evaluations. Given the average number of evaluations was 18 (median = 19), evaluators were categorized as a) conducting fewer evaluations ($3 – 18; n = 9$); and b) conducting a higher proportion ($19 – 39; n = 11$). While not significant, the results suggested those who conducted more evaluations were more likely to provide a 'non-consent' specifier $X^2 (2, N = 20) = 5.45, p = 0.07$. Notably, these analyses were re-ran categorizing the evaluator into having ever provided a 'non-consent' specifier ($n = 15$) or not ($n = 5$). Similar findings were found for evaluator degree, gender, and proportion of evaluations conducted (see Tables 7 - 9).

In an attempt to understand when evaluators provided the 'non-consent' specifier, victim age was explored. A chi-square test was performed to examine the relationship between victim age and whether a 'non-consent' specifier was used. Specifically, this relationship was explored both between the victim age of the offender's index offense, as well as the offender's 'preferred' victim age (see Tables $12 – 14$). Given that perpetrators of victims ten years and younger never received a 'non-consent' specifier, this age group was not included in the analyses. Results suggested the use of the 'non-consent' specifier was not dependent on victim age for his index offense [$X^2 (3, N = 276) = 7.67, p = 0.53$]. However, the use of the 'non-consent' specifier

27

demonstrated significant results as it varied by age group [X^2 (3, $N = 358$) = 10.75, $p = 0.01$][5] for

the offender's *preferred* victim age (see Table 14). Notably, there was not a significant

difference in the application of the 'non-consent' specifier between those offenders whose

preferred victim age was an adult compared to those who had a history of offending against

multiple age groups [X^2 (1, $N = 279$) = 0.55, $p = 0.46$].

Hebephilia.

The term 'hebephilia' was used far less frequently compared to 'non-consent.' As well,

there was less consistency in its use. To be inclusive, the 'hebephilia combined' variable was

explored. At noted, hebephilia terminology was used 16 times out of the 107 times OSPD was

diagnosed (15%; see Table 3). Again, it appears as if 'hebephilia' was used, on average, in 15%

of the OSPD diagnoses. However, three evaluators used this term half the time they provided an

OSPD diagnosis, 9 evaluators *never* used this specifier, and eight evaluators used 'hebephilia' as

a specifier 8 – 30% of the time they diagnosed OSPD (see Table 6).

First evaluator differences in the use of this term (see Tables 10 - 12) were explored.

Given the low frequency in use of 'hebephilia,' evaluators were categorized as 'never' using this

term ($n = 11$) or using this term 'once or more' ($n = 9$). As with non-consent, the degree of the

psychologist did not influence the use of 'hebephilia' [X^2 (2, $N = 20$) = 0.64, $p = 0.2$]. However,

evaluator gender demonstrated significant results [X^2 (2, $N = 20$) = 4.10, $p = 0.04$]; suggesting

males were more likely to *never* use this label. Additionally, the results suggested those who

conduct more evaluations are more likely to use the term 'hebephilia' [X^2 (1, $N = 20$) = 7.10, $p =$

0.00].

---

[5] These analyses were conducted based on all age groups as identified in the Methods section, and then combined age groups (10 and under; 12-16). Results did not vary and only the combined age group analyses are presented.

Chi-square analyses were conducted to explore the relationship between victim age and the offender receiving a 'hebephilia' specifier (see Tables 15 – 16; see also, footnote 3). Results suggested the use of 'hebephilia' was dependent on victim age for the index offense [X^2 (5, $N = 306) = 11.44$, $p = 0.04$]. Specifically, of the 13 individuals who received a 'hebephilia' specifier, one had offended against a victim 10 and under; two offended against a victim 12 years of age; two offended against a victim in the 13 – 14 age group; two offended against victims in the 15 – 16 age group; three offended against victims of adult age; and three had offended against victims in multiple age groups (three cases were missing the index victim age). On the other hand, the use of 'hebephilia' did not significantly vary between age group of the offender's *preferred* victim age. The majority of cases ($n = 15$) who received the 'hebephilia' specifier offended against victims in multiple age groups; whereas, one offender who received a 'hebephilia' specifier offended against children 10 years or younger.

CHAPTER 4: DISCUSSION

Previous research (e.g., Becker et al., 2003; Elwood et al., 2010; Jackson & Richards, 2007; Janus & Walbek, 2000; Levenson, 2004; Lu et al., 205; McLawsen et al., 2012) has found paraphilia NOS to be one of the most commonly diagnosed paraphilic disorders within the SVP context. However, these findings were based on an earlier version of the DSM and not the DSM-5, which modified the paraphilia NOS category by dividing it into two separate diagnostic categories—OSPD and UPD. Additionally, this revision now requires clinicians to specify an OSPD diagnosis by indicating the source of sexual arousal for the individual being evaluated. Until now, however, no study had systematically explored how the revision to paraphilia NOS has affected diagnostic tendencies within SVP evaluations. Findings from this study suggest that OSPD has "replaced" paraphilia NOS as it was the second most common paraphilic diagnosis proffered in these SVP evaluations after pedophilia. While the aim of the OSPD label is presumably to provide better clarity about paraphilic interest, results from this study suggest that diagnostic reliability for OSPD did not improve above the former, arguably more ambiguous, paraphilia NOS diagnosis in DSM-IV (see, Packard & Levenson, 2006; Perillo et al., 2014). Further, while 'non-consent' and 'hebephilia' were the most frequently-used labels to specify an offender's source of arousal, this study demonstrated that there is little consistency in the specifiers evaluators are using; and that there does not appear to be a standard, methodological approach for how SVP evaluators should determine an appropriate specifier.

Aim 1. Prevalence of Diagnoses

The first aim of this study was to explore the prevalence of diagnoses used in SVP evaluations since the publication of the updated DSM. In the current study, OSPD was the second most commonly diagnosed paraphilic disorder (28%), with pedophilia being the most

common (30%). UPD was the third most common diagnosis (17%), which was diagnosed in higher frequency than the rest of the paraphilic disorders (e.g., exhibitionism, sexual sadism).

Aim 2. Use of OSPD Specifiers

With the newest edition of the DSM, clinicians who diagnose OSPD must communicate the specific reason the clinical presentation does not meet the criteria for one of the listed paraphilic disorders (APA, 2013). Thus, the second aim of this study was to explore which specifiers were used with an OSPD diagnosis.

In the present study evaluators did provide some type of label a majority (85%) of the time they diagnosed OSPD. That said, there was 16 evaluations (15% of the time), in which clinicians proffered this diagnosis and did not communicate a specifier. That is, they provided an OSPD diagnosis but provided no indication as to why the offender did not meet criteria for one of the other established paraphilic disorders or what the source of his sexual arousal was. It is not known why this occurred. Further exploration revealed one evaluator never provided a label. However, there were nine evaluators who typically provided a label, but on one or two occasions did not. Given that the DSM-5 specifies that a label must be provided, there is not a clear explanation for why evaluators are inconsistently doing so.

It was hypothesized the specifiers 'non-consent' and 'hebephilia' would be the most frequently used terms. This hypothesis was partly supported. The distinct term 'non-consent' was the most frequently used specifier; it was provided 57% of the time OSPD was diagnosed. This is in line with previous research that demonstrated, of those civilly committed sexual offenders diagnosed with paraphilia NOS, 57% - 67% of them received a 'non-consent' label (Elwood et al., 2010; Jackson & Richards, 2007; McLawsen et al., 2012). While the distinct term 'hebephilia' was used with far less frequency (7%), it was the next most commonly provided label.

While 'non-consent' and 'hebephilia' were the most common labels, there was a variety of terminology used, and there does not appear to be consistency in the language used for these specifiers. On the one hand, a majority of the terms could be conceived to mean 'non-consent' or 'hebephilia.' For example, the term 'biastophilia,' which is often interchanged for paraphilia non-consent, was used with some frequency in the current sample. Similarly, the construct of hebephilia appeared to be communicated with labels such as 'sexually attracted to teenagers.' However, when labels such as 'sexually attracted to teenagers' are used, it is hard to discriminate which age group the clinician may be referring to. While 13- and 14-year-olds may be considered 'teenagers,' and could fall within most definitions of hebephilia, 15 - 17-year-olds are still considered 'teenagers,' but are more likely to physically resemble advanced Tanner stages (Tanner, 1990)—with physical features more characteristic of adults rather than pubescent youth. This is problematic because—although an agreed upon definition for hebephilia is lacking—the term has generally been defined as a sexual interest for pubescent-aged adolescents—those whom physical characteristics resemble the development in Tanner stages 2 – 3 (e.g., 11-14 years old; Stephens et al., 2016). Further, there were times in which the labels clinicians used appeared to combine both constructs (e.g., 'non-consensual sexual activity with an adolescent').

Aim 3. Consistency in Evaluators' Diagnostic Tendencies

Despite OSPD being the second most commonly proffered diagnosis, there was little consistency in evaluators agreeing on the presence of this diagnosis. For one, there was a wide range in the frequency with which evaluators used this diagnosis. On the one hand, two (9%) clinicians proffered this diagnosis in 50% or more of the evaluations they conducted and five evaluators (24%) provided it more than a third of the time; However, eight (38%) of the clinicians provided this diagnosis in less than a quarter of the evaluations they conducted, and one clinician never used this diagnosis in any of the evaluations he or she conducted. Secondly,

there appeared to be little agreement among evaluators in diagnosing OSPD. Evaluators agreed on the presence of this disorder less than 50% of the time this disorder was diagnosed.

It is important to note much debate stemmed from the use of paraphilia NOS within SVP evaluations—criticizing this diagnosis as a residual category which lacks reliability and validity (e.g., Frances & First, 2011b; Miller et al., 2003; Polaschek, 2003). While OSPD intended to increase diagnostic clarity by specifying one's source of sexual arousal, results from this study suggest inter-rater reliability continues to be as poor as the original paraphilia NOS diagnosis (see, for example, Perillo et al., 2014). In fact, previous studies demonstrated slightly better inter-rater reliability of paraphilia NOS than did OSPD in the current study (see Table 4; Packard & Levenson, 2006; Perillo et al., 2014). That said, when a paraphilia NOS diagnosis was computed in the current study—by combining diagnoses of OSPD and UPD—reliability results improved slightly (see Table 4).

In contrast, pedophilia—a paraphilic disorder with more explicit diagnostic criteria than OSPD—demonstrated greater agreement. For one, pedophilia was used by all the evaluators. Specifically, evaluators offered this diagnosis in 13% to 47% of the evaluations they conducted. Additionally, evaluators tended to agree on the presence or absence of pedophilia in the offender they were evaluating (PA= 91%). From these findings one may conclude that diagnoses with more explicit diagnostic criteria improves inter-rater reliability, and perhaps increases the likelihood an evaluator is willing to proffer a paraphilic diagnosis.

Reliability of specifiers.

Despite 'non-consent' (i.e., non-consent combined) being used with high prevalence in the current study, evaluator's frequency in use of this term ranged greatly. One evaluator used this label every time he or she diagnosed OSPD; over half of the evaluators used this specifier more than 50% of the time they diagnosed OSPD; still, a quarter of the evaluators never used this

specific term when diagnosing OSPD. Evaluators agreed in providing this label about a third of the time (35%).

Similarly, the use of 'hebephilia' (i.e., hebephilia combined) among evaluators also ranged. Three evaluators used the label 'hebephilia' half the time they provided an OSPD diagnosis; however, nine of the evaluators never once used this label. Evaluators agreed in providing this label less than a third of the time (27%).

Aim 4. Understanding Differences between Evaluators

The interpretation of these findings is limited as it cannot be known if an individual genuinely met diagnostic criteria of any paraphilic diagnosis. However, these findings still have important implications. For one, consistent with previous research demonstrating poor inter-rater reliability for paraphilia NOS (e.g., Perillo et al., 2014), and as hypothesized, there continues to be little agreement between evaluators when diagnosing OSPD or UPD. Further, there is even less consistency amongst evaluators in the specifiers they are using.

Guarnara, Murrie, and Boccaccini (2017) discuss several factors that may influence the reliability of forensic evaluators' opinions. One such explanation is differences in education and specialized training. The only available data relevant to this factor in this study was evaluator degree, (e.g., Ph.D. versus Psy.D.), which did not influence whether one was more likely to use a 'non-consent' or 'hebephilia' specifier. That said, the quality or specialization of training one received cannot be assessed based on degree alone.

Adversarial allegiance is another factor Guarnera and colleagues (2017) identify as a potential source for unreliability. Previous research has elucidated a tendency for forensic evaluators to reach conclusions favorable to the side which retained them (Murrie & Boccaccini, 2015; Murrie et al., 2009). The current sample, however, could theoretically be considered 'neutral'. Specifically, the evaluators in the current study were comprised of psychologists who

34

are contracted by the Florida Department of Children and Families (DCF) to assist in making a determination as to whether the state should petition a commitment hearing. If a commitment hearing is held, the defense may request an independent psychological evaluation. The prosecution (i.e., the State of Florida) may also request an independent psychological evaluation, or simply proceed with the findings from the SVPP contracted evaluators. Therefore, given that these evaluators are contracted by DCF, and not the defense or the prosecution, the influence of adversarial allegiance should—at least theoretically—be obsolete.

Nevertheless, there may be implicit ideology associating Florida DCF with the prosecuting body—that being, the State of Florida. As such, a clinician may be implicitly motivated to find an opinion favorable to civil commitment as that tends to be the aim of the prosecution. In order to civilly commit an individual, one must present with a mental abnormality that places him or her at risk of committing a sexually violent offense. Therefore, the prosecution would seek a mental diagnosis to fit this criterion. Interestingly, there was a trend in the data to suggest that those evaluators who conducted a higher frequency of SVP evaluations were more likely to use labels such as 'non-consent' and 'hebephilia.' Keeping with adversarial allegiance, one could suggest those evaluators who are more consistently retained by the state, or whom wish to be consistently retained by the state, may provide diagnoses favoring civil commitment. That said, the data of the present study in no way prove this was the practice. Additionally, the present sample of offenders may have been at an increased risk for sexual recidivism, compared to the larger pool of offenders evaluated under SVP legislation; this is explained in greater detail below.

Alternatively, another source of bias may have influenced the findings. Improved consistency in diagnostic tendencies was demonstrated when evaluators agreed on commitment decisions; and further when they both recommended the offender for civil commitment. While

the second evaluator is not formally notified a first evaluation occurred, it is possible to find out anecdotally (e.g., the offender self-reports). Given that the Florida SVPP process requires a second evaluation take place if the first evaluator recommends civil commitment, the second evaluator may—consciously or not—seek to also support a civil commitment recommendation. Indeed, in the current sample evaluators were more likely to agree on commitment decisions than disagree; and consistency in diagnoses rendered improved when the evaluators both recommended commitment. On the one hand, these cases may have been more straightforward, and "easier" to satisfy SVP commitment criteria. On the other hand, it is possible the second evaluator may have been implicitly motivated to "find" a diagnosis. If the offender did not meet criteria of the more explicitly defined paraphilic disorders (e.g., pedophilia;), then perhaps he was diagnosed with OSPD.

Aside from these biases, evaluators' individual differences can also influence one's approach to forensic assessment and contribute to unreliability (Guarena et al., 2017; Miller et al., 2011). Further, preexisting values and beliefs may also influence one's approach to conducting a forensic evaluation, diagnostic tendencies, and opinions rendered (Neal & Brodsky, 2016). However, data on the evaluators' personality characteristics, attitudinal perspectives, values, and morals was not available in this field study.

Implications for Research

Future research may seek a more experimental approach to explore how individual differences may influence one's approach to diagnostic tendencies within SVP evaluations. For example, previous research has explored the association between risk assessment scoring tendencies and personality traits of the evaluator (see Miller et al., 2011); another study has explored awareness of bias in clinical work by conducting focus groups with forensic psychologists (see Neal & Brodsky, 2016).

Focus group methodology could also be beneficial in exploring how evaluators come to the decisions they do and the real-world complications SVP evaluators face in these assessments. For example, some defendants opt-out of an in-person interview, making already difficult diagnostic decision-making even more tough. Multi-site focus groups may help to begin a conversation eliciting the practical obstacles SVP evaluators are challenged with and inform how diagnostic decisions are made.

Still, additional field studies are also necessary. While the current study was inclusive of all the evaluations conducted by the Florida SVPP since the DSM-5 was published (as long as both evaluators used the DSM-5), several limitations to this dataset exist. For one, there were only 21 distinct evaluators and it was solely evaluations conducted in Florida. As such, the findings from this study may not be reflective of the practice of SVP evaluations across the twenty jurisdictions which apply this legislation. Additionally, given an aim of this study was to explore inter-rater reliability, this sample was limited to those offenders who received two SVP evaluations. This sample may pose distinct characteristics that can limit generalizing the current findings to those offenders who receive only one evaluation. For example, the practice of the Florida SVPP requires defendants to receive an additional evaluation if his first evaluation rendered a civil commitment recommendation. As such, the defendants in the current sample may be considered of a higher risk, or greater diagnostic complexity, in comparison to those who only received one evaluation. Similarly, interpretation of the diagnostic prevalence rates should be considered within the context of these evaluations. As discussed, the sample of evaluators in the current study were those contracted by Florida DCF. As such, this sample did not include any defense-retained SVP evaluators who may be—consciously or subconsciously—less likely to find a paraphilic diagnosis. Further research might examine whether there are patterns in how defense or prosecution retained evaluators apply paraphilia diagnoses in the SVP context.

Implications for Policy

On the one hand, this study has contributed to the field as it was the first to provide some transparency to the type and frequency of specifiers applied in SVP evaluations. On the other hand, the results of this study may have furthered muddled the perspective about the practice of SVP evaluations. While this study showed OSPD being diagnosed in high prevalence; the results also suggest little reliability among evaluators, with evaluators using an idiosyncratic approach to providing OSPD specifiers.

Although 'non-consent' and 'hebephilia' were the most frequently used specifiers, there does not appear to be a clear, shared definition of the specifiers that are being used. For example, 'hebephilia' was used in cases with victims across a range of ages—even for defendants who's index offense included an adult-aged victim. Likewise, labels often overlapped (e.g., 'non-consensual activity with an adolescent'). Additionally, there was great variety in labels outside of the specific 'non-consent' and 'hebephilia' terminology. As discussed above, many of the other reported labels might relate to the constructs of paraphilia non-consent (e.g., biastophilia), or hebephilia (e.g., sexual attraction to teenager); however, making such assumptions within a litigious context is haphazard and unsystematic. Further, there lacks reliability between evaluators at reaching an OSPD diagnosis—in general—as well as a lack of consistency between evaluators at providing the same specifier.

On the other hand, from a qualitative perspective, evaluators appear to demonstrate greater consistency than quantitative results suggest. Despite less than chance agreement on an OSPD diagnosis, evaluators, overall, seem to agree on the presence of deviant sexual arousal. For example, of the 61 times that evaluators did not agree on the presence of an OSPD diagnosis, the opposing evaluator provided a paraphilic diagnosis of some type 61% (n = 37) of the time.

Typically, this was a diagnosis of UPD (51%, $n = 19$), but sometimes was a diagnosis of pedophilia (32%, $n = 12$) or sexual sadism (14%, $n = 5$).

Taken together, these findings suggest the need for a standardized approach to SVP evaluations. For one, there needs to be consensus in the field about the use of OSPD, in general, and non-consent and hebephilia more specifically; this is discussed in greater detail below. However, there does not appear to be consistency between the states about how SVP evaluators are trained or approach such assessments.

Forensic evaluations are inherently difficult (Guarnera et al., 2017). SVP evaluations are not 'alone' in demonstrating low levels of consistency between evaluators. For example, a recent meta-analysis demonstrated fair to moderate levels of agreement between evaluators in adjudicative competency and sanity evaluations (Guarnera & Murrie, in press). As such, the inconsistency between evaluators posed in this study is not necessarily at fault of the evaluators; forensic mental health assessment—overall—is not perfect and unlikely ever will be (see, Mossman, 2013). That said, research has shown that more rigorous state-level training and certification can improve the field reliability of forensic opinions (Gowensmith, Sledd, & Sessarego, 2014). The results from this study suggest that, despite low levels of inter-rater reliability, evaluators are—more often than not—still on "the same page" in terms of diagnostic ideology. As such, it may be advantageous for states with SVP legislation to mandate training focused on a standardized method of diagnosing OSPD and guidelines for determining appropriate specifiers.

Application to the courtroom.

Although King and colleagues (2014) found that a diagnosis of paraphilia non-consent was always admitted into evidence, more recently, there have been some Frye hearings which ruled paraphilia non-consent, and hebephilia as inadmissible for SVP commitment (e.g., *Matter*

of State of New York v. Ralph P., 2016; and *Matter of State of New York v. Jason C.,* 2016) following testimony by experts who opined that paraphilia non-consent and hebephilia lack empirical support for their reliability and validity. Nevertheless, the court's decisions in these matters were based on the opinion that there is not a general consensus in the field on the acceptability of these constructs as diagnoses. At first glance the findings from this study suggest there is fairly regular use of these constructs within SVP evaluations. That said, there were important differences between evaluators in their practice of these specifiers; whereas some evaluators used these specifiers for a majority of their OSPD diagnoses, others never used either of these specifiers. As such, a mere 'hand-count' of who uses these specifiers may not be an accurate representation of scientifically-sound practice as some clinicians seem to more regularly use these specifiers than do others. Moreover, regular use does not necessarily imply correct use—what clinicians are doing in the field may not always be what they 'should' be doing based on 'good science' and ethical guidelines. Further, the majority of states have a higher standard of admissibility of evidence (i.e., Daubert criteria) which requires testimony to be based on scientifically valid reasoning and assessment.

When psychologists are conducting these evaluations it is likely there are some individuals whom they consider to be a sexually violent risk, but ultimately can not recommend for civil commitment because without a nexus between a mental illness and sexually violent behavior, there is no foundation for such preventative detention. Thus, recommendation for civil commitment must be based by some scientific validity rather than a "simpl[e] political choice of those in power" (e.g., a judge; Prentky, Janus, Barbaree, Schwartz & Kafka, 2006; p. 361).

While the impetus is on the psychologist to practice in a scientifically-valid manner; it is ultimately up to the decision maker(s) to evaluate the science proffered in the courtroom and determine the weight to give the expert's testimony (Janus & Prentky, 2003). Problematically,

however, judges are not always equipped to recognize 'good science' and the pressure inherent in the SVP process increases the chance of 'bad science' being introduced into the courtroom (Prentky et al., 2006). Testimony about empirically-validated mental disorders can have a positive impact within the SVP courtroom by providing further clarity between how science can inform the law; as well as balancing consequential decisions between public safety and impeding on constitutional rights. That said, unreliable and inconsistent use of diagnoses can cause legal actors to lose faith in the value of scientific testimony from the field of psychology and ultimately undermine the constitutionality of SVP civil commitment. As such, diagnoses rendered to satisfy SVP commitment should be empirically validated.

Reliability and validity of specifiers.

As of now, the field lacks research on the etiology and pathology of OSPD in general, and paraphilia non-consent and hebephilia, more specifically. Along these lines, there is a lack of research—for both non-consent and hebephilia—to support the *Hendricks* standard which requires that the mental abnormality (i.e., non-consent and hebephilia) be linked to making it "difficult or impossible to control dangerous behavior" (*Kansas v. Hendricks,* 1997). That is, there is a dearth of research to demonstrate the predictive validity of non-consent or hebephilia and future sexually violent behavior. Likewise, there is a lack of research to support the *Crane* standard which requires that the nature and severity of the mental disorder "be sufficient to distinguish the dangerous sexual offender whose serious mental illness, abnormality, or disorder subjects him to civil commitment from the dangerous but typical recidivist convicted in an ordinary criminal case" (*Kansas v. Crane,* 2002). As Miller and colleagues (2005) opined, there is a paucity of empirical support to suggest paraphilia non-consent could reliably distinguish a class of rapists who have a mental abnormality from those who do not. Other scholars have expressed similar arguments for both non-consent and hebephilia (see Footnote 1). Implications

41

for additional research to explore the prevalence of OSPD and its associated specifiers was addressed above. Research exploring the etiology, pathology, and prevalence of other potential paraphilias (e.g., non-consent and hebephilia), as well as research exploring their relation to risk and recidivism, is also warranted.

Given that diagnostic decisions represent (almost exclusively) the work of psychologists, the American Psychological Association (APA) or the Association for the Treatment of Sexual Abusers (ATSA) should evaluate findings from this, and future studies to formulate an opinion that represents the general professional consensus on use of these constructs within SVP evaluations. The use of unreliable diagnoses can have significant implications across the board. The practice of 'bad science' can hamper the credibility of psychology within the legal system— a status the field has worked hard to achieve. Similarly at stake when 'bad science' is used, is our communities' safety and financial expenditure. Further, and arguably most important, the use of inadequate science can lead to profound consequences for the offender by violating his or her civil liberties with an unconstitutional commitment.

Psychologists serving in the role as an SVP evaluator should abide by ethical guidelines recommending that practice is guided by a scientific foundation—utilizing reliable and valid principles and methods. It is important clinicians evaluate all evidence and consider how it may contradict his or her diagnostic opinion. Psychologists should provide the data and reasoning he or she used to reach a diagnostic conclusion, how such a diagnosis is relevant to risk, and clearly communicate the limits posed in the evaluation. Given ethical guidelines which call for objectivity and transparency in forensic assessment (American Psychological Association, 2013), as well as the recent research on adversarial allegiance (Murrie & Boccaccini, 2015; Murrie et al., 2009), clinicians should be willing to subject their diagnostic opinion and reasoning to

critique by colleagues and advisors (King et al., 2014) to enhance reliability and validity of his or her decision and assess any bias posed in diagnostic decision-making.

Conclusion

In all, it is the hope these results and suggestions are used to advance 'good science' for SVP legislation. Political agendas aimed at protecting communities and reducing sexual violence can inadvertently mask poor clinical practice, and judges are not trained to evaluate clinical decision making. Given the profound consequences SVP legislation has on multiple parties, it is imperative for researchers to advance our understanding of the etiology and pathology of paraphilias and for clinicians to adhere to empirically-supported principals and methodology.
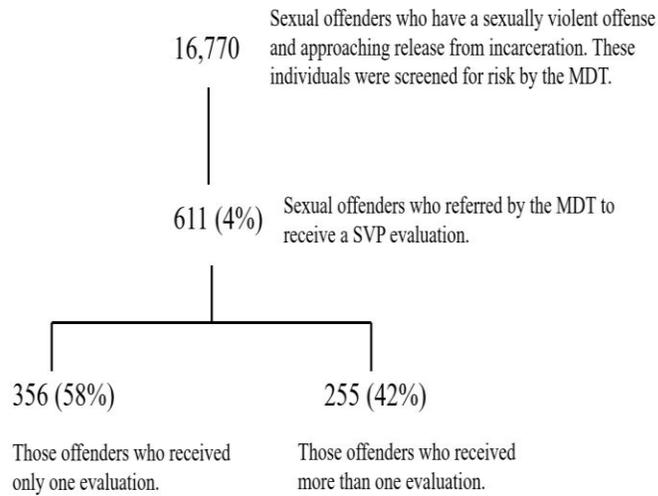
Figure 1. Sexual offenders who received a SVP evaluation between May, 2013 and June, 2017.

Table 1

Prevalence of Diagnoses

Diagnosis	Prevalence % (*N*)
Any Paraphilia	**77.89% (296)**
Pedophilia	30.0% (114)
OSPD	28.42% (107)
UPD	17.37% (66)
Exhibitionism	5.79% (22)
Sexual Sadism	3.16% (12)
Voyeurism	3.16% (12)

Note. OSPD = Other specified paraphilic disorder; UPD = Unspecified paraphilic disorder; ASPD = Antisocial personality disorder

Table 2
Coding Scheme and Specifiers used in OSPD Diagnoses

	Non-Consent	Non-Consent Combined	Hebephilia	Hebephilia Combined	Other Labels
	Non-Consent (42)	Non-Consent (42)	Hebephilia (5)	Hebephilia (5)	Sadistic Features (2)
	Non-Consenting Sex	Non-Consenting Sex	Ephebophilia, Hebephilia	Ephebophilia, Hebephilia	Exhibitionistic and Sadistic Features
	Nonconsensual Sex	Nonconsensual Sex		Pedo-Hebephilia	Bestiality
	Nonconsenting; in a Controlled Environment	Nonconsenting; in a Controlled Environment		with Exhibitionistic and probable Pedophilic/ Hebephilic Features	Zoophilia
	Nonconsent, in a Controlled Environment (2)	Nonconsent, in a Controlled Environment (2)		Sexually Attracted to Teenagers, in Remission in a Controlled Environment	Sexting
	Non-Consenting Partners	Non-Consenting Partners		Sexually attracted to Teenagers, Males and Females, Nonexclusive Type	NOS
	Nonconsenting Persons	Nonconsenting Persons		Adolescent Females	in a Controlled Environment (2)
	Non-Consenting Partner (NOS)	Non-Consenting Partner (NOS)		Attraction to Adolescent Females	
	Nonconsenting with Females	Nonconsenting with Females		Sexually Attracted to Young Pubescent Females, in a Controlled Environment	
	Non-Consenting Sex with Females, in Remission in a Controlled Environment	Non-Consenting Sex with Females, in Remission in a Controlled Environment			
	with Arousal to Non-Consensual Sex	with Arousal to Non-Consensual Sex			
	features of Non-Consenting Persons and Exhibitionism	features of Non-Consenting Persons and Exhibitionism			
	Nonconsent, with Sadistic Features	Nonconsent, with Sadistic Features			

	Nonconsent, Sadistic and Exhibitionistic Features	Nonconsent, Sadistic and Exhibitionistic Features			
		Biastophilia (4)			
		with features of Hebephilia and Biastophilia		with features of Hebephilia and Biastophilia	
		Biastophilia/ Paraphilic Rape			
		with features of Paraphilic Coercion and Courtship Disorder			
	Complex: Nonconsent, Force, Violence, Compulsive Use of Pornography and Telephone Scatologia	Complex: Nonconsent, Force, Violence, Compulsive Use of Pornography and Telephone Scatologia			
	Necrophilia Nonconsenting	Necrophilia Nonconsenting			
	Nonconsent; Hebephilia	Nonconsent; Hebephilia	Nonconsent; Hebephilia	Nonconsent; Hebephilia	
	Non-Consensual Sexual Activity with Adolescent	Non-Consensual Sexual Activity with Adolescent		Non-Consensual Sexual Activity with Adolescent	
	Nonconsensual Sexual Activity with Age Inappropriate Individuals	Nonconsensual Sexual Activity with Age Inappropriate Individuals			
Total	**61**	**68**	**7**	**16**	**9**

Note. Specifiers were grouped based on the category they were most closely related to. Numbers in parentheses indicate the number of times the label was used, if used more than once.

Table 3

Prevalence of 'Hebephilia' and 'Non-consent' Specifiers

Specifier	Total Evaluations (N = 380)		Total OSPD Diagnoses (N = 107)	
	% either evaluator	% both evaluators	% either evaluator	% both evaluators
Hebephilia	1.84%	0.26%	6.54%	0.93%
Hebephilia Combined	4.21%	0.52%	14.95%	1.87%
Non-consent	15.79%	2.37%	56.01%	8.41%
Non-consent Combined	17.89%	3.16%	63.55%	11.21%

Table 4

Diagnostic Reliability across Evaluators

Diagnosis	Kappa (95% CI)	PPV (95% CI)	NPV (95%CI)	PA (95% CI)	PA+ (95% CI)	PA- (95% CI)	n
Any Paraphilia	.23** (.07 - .67)	.90 (.84 - .95)	.29 (.18 - .43)	.73 (.66 - .79)	.82 (.78 - .87)	.38 (.25 - .51)	190
Pedophilia	.78** (.66 - .89)	.91 (.81 - .97)	.90 (.84 - .95)	.91 (.85 - .94)	.85 (.79 - .92)	.93 (.90 - .96)	190
Exhibitionism	.58*** (.26 - .88)	.60 (.26 - .81)	.98 (.94 - .99)	.96 (.91 - .98)	.60 (.35 - .85)	.98 (.96 - .99)	190
Sexual Sadism	-0.04 (-.04 - .31)	.00 (.00 - .45)	.94 (.95 - .96)	.92 (.87 - .95)	.00 (.00 - .00)	.96 (.93 - .98)	190
Voyeurism	.53** (.14 - .83)	.50 (.12 - .88)	.99 (.96 - 1.00)	.97 (.94 - .99)	.55 (.19 - .90)	.99 (.97 - 1.00)	190
OSPD	.21**(.07 - .36)	.48 (.33 - .63)	.75 (.67 - .82)	.68 (.61 - .74)	.43 (.31 - .55)	.78 (.72 - .83)	190
UPD	.16* (-.01 - .32)	.30 (.16 - .49)	.85 (.79 - .90)	.76 (.69 - .82)	.30 (.16 - .45)	.85 (.81 - .90)	190
Non-Consent	.17* (.00 - .34)	.36 (.18 - .57)	.84 (.77 - .89)	.78 (.71 - .84)	.30 (.15 - .45)	.87 (.83 - .91)	190
Non-Consent Combined	.22** (.05 - .38)	.44 (.25 - .65)	.82 (.75 - .88)	.77 (.70 - .83)	.35 (.21 - .50)	.86 (.82 - .90)	190
Hebephilia	.27** (-.02 - .71)	.20 (.01 - .72)	.99 (.97 - 1.00)	.97 (.94 - .99)	.29 (-.15 - .72)	.99 (.97 - 1.00)	190
Hebephilia Combined	.22*** (-.07 - .51)	.20 (.03 - .56)	.98 (.94 - .99)	.94 (.90 - .97)	.27 (-.03 - .56)	.97 (.95 - .99)	190
Paraphilia NOS^	.27*** (.54 - .72)	.64 (.53 - .75)	.63 (.54 - .72)	.64 (.57 - .71)	.60 (.52 - .69)	.67 (.60 - .75)	190
Packard & Levenson (2006) Paraphilia NOS	.36*** (.25 - .47)	.65 (.58 - .71)	.71 (.66 - .75)	.68 (.63 - .74)	.47 (.39 - .55)	.56 (.49 - .63)	277
Perillo et al (2014) Paraphilia NOS	.35* (.20 - .50)	.52 (.37 - .68)	.85 (.78 - .89)				375

Note: The Bloom, Fischer, and Orme (1999) standard for kappa agreement was used ("poor" = below 0.60, "fair" = 0.60 – 0.74, and "good" = 0.75 and above) to assess for poor to good reliability. PA = Proportion of agreement, overall; + PA = Proportion of agreement diagnosis is present; - PA = Proportion of agreement diagnosis is not present.

^Paraphilia NOS was created from combining OSPD and UPD.

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

Table 5

Diagnostic Reliability across Evaluators based on Commitment Decisions

Diagnosis	Kappa (95% CI)	PPV (95% CI)	NPV (95%CI)	PA (95% CI)	PA + (95% CI)	PA – (95% CI)	N
Any Paraphilia							
Full Study	.23*** (.07 - .67)	.90 (.84 - .95)	.29 (.18 - .43)	.73 (.66 - .79)	.82 (.78 - .87)	.38 (.25 - .51)	190
Agree on Commit	.59*** (0.39 - .78)	.98 (.92 - 1.00)	.52 (.31 - .72)	.88 (.81 - .93)	.93 (.89 - .97)	.65 (.48 - .82)	120
Do Not Agree on Commit	-0.18 (-.36 - .01)	.72 (.55 - .85)	.10 (.02 - .27)	.45 (.33 - .57)	.60 (.48 - .71)	.14 (.00 - .27)	69
Recommended Commit	.10 (-.17 - .38)	.98 (.92 - 1.00)	.10 (.00 - .45)	.89 (.81 - .94)	.94 (.91 - .98)	.15 (-.11 - .42)	100
OSPD							
Full Study	.21** (.07 - .36)	.48 (.33 - .63)	.75 (.67 - .82)	.68 (.61 - .74)	.43 (.31 - .55)	.78 (.72 - .83)	190
Agree on Commit	.29** (.11 - .47)	.54 (.37 - .71)	.76 (.65 - .85)	.69 (.60 - .77)	.52 (.38 - .66)	.77 (.70 - .84)	120
Do Not Agree on Commit	.04 (-.17 - .24)	.27 (.06 - .61)	.72 (.59 - .83)	.65 (.53 - .76)	.20 (.01 - .39)	.78 (.69 - .86)	69
Recommended Commit	.28** (.08 - .47)	.53 (.35 - .70)	.75 (.63 - .85)	.67 (.57 - .76)	.54 (.39 - .68)	.74 (.66 - .83)	100
Non-consent							
Full Study	.17* (.00 - .34)	.36 (.18 - .57)	.84 (.77 - .89)	.78 (.71 - .84)	.30 (.15 - .45)	.87 (.83 - .91)	190
Agree on Commit	.19** (-.01 - .40)	.38 (.18 - .62)	.82 (.73 - .89)	.75 (.66 - .82)	.35 (.17 - .52)	.85 (.79 - .90)	120
Do Not Agree on Commit	.08 (-.17 - .33)	.25 (.01 - .81)	.86 (.75 - .93)	.83 (.72 - .91)	.14 (-.11 - .39)	.90 (.85 - .96)	69
Recommended Commit	.17 (-.05 - .38)	.38 (.18 - .62)	.78 (.68 - .87)	.71 (.61 - .80)	.36 (.18 - .53)	.81 (.75 - .88)	100
Non-consent Combined							
Full Study	.22** (.05 - .38)	.44 (.25 - .65)	.82 (.75 - .88)	.77 (.70 - 0.83)	.35 (.21 - .50)	.86 (.82 - .90)	190
Agree on Commit	.26** (.06 - .46)	.48 (.27 - .69)	.81 (.72 - .89)	.75 (.66 - 0.82)	.42 (.25 - .59)	.84 (.78 - .90)	120
Do Not Agree on Commit	.06 (-.16 - .27)	.25 (.01 - .81)	.83 (.72 - .91)	.80 (.68 - 0.88)	.13 (-.10 - .35)	.89 (.83 - .94)	69
Recommended Commit	.24* (.03 - .45)	.48 (.27 - .69)	.78 (.67 - .87)	.71 (.61 - 0.80)	.43 (.26 - .60)	.81 (.74 - .87)	100

Hebephilia

Full Study	.27** (-.02 - .71)	.20 (.01 - .72)	.99 (.97 - 1.00)	.97 (.94 - 0.99)	.29 (-.15 - .72)	.99 (.97 - 1.00)	190
Agree on Commit	.39*** (-.16 - .94)	.33 (.01 - .91)	.99 (.94 - 1.00)	.98 (.93 - 0.99)	.40 (-.14 - .94)	.99 (.97 - 1.00)	120
Do Not Agree on Commit	.00 (.00 - .00)	n/a	.97 (.90 - 1.00)	.97 (.90 - 1.00)	.00 (.00 - .00)	.99 (.96 - 1.01)	69
Recommended Commit	.39*** (-.17 - .94)	.33 (.01 - .91)	.98 (.93 - 1.00)	.97 (.91 - 0.99)	.40 (-.14 - .94)	.98 (.97 - 1.00)	100
Hebephilia Combined							
Full Study	.22** (-.07 - .51)	.20 (.03 - .56)	.98 (.94 - .99)	.94 (.90 - 0.97)	.27 (-.03 - .56)	.97 (.95 - 0.99)	190
Agree on Commit	.34*** (-.04 - .71)	.29 (.04 - .71)	.98 (.94 - 1.00)	.94 (.89 - 0.98)	.36 (.00 - .73)	.97 (.95 - 0.99)	120
Do Not Agree on Commit	-.04 (-.07 - .00)	n/a	.96 (.87 - .99)	.93 (.84 - 0.98)	.00 (.00 - .00)	.96 (.93 - 1.00)	69
Recommended Commit	.26* (-.18 - .70)	.17 (.01 - .64)	.99 (.94 - 1.00)	.95 (.89 - 0.98)	.29 (-.15 - .72)	.97 (.95 - 1.00)	100

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 6

Evaluator's Use of 'Non-consent' and 'Hebephilia' Specifiers

Evaluator	# Evaluations	% OSPD	% Non-Consent	% Hebephilia
A	39	31%	83%	8%
B	28	7%	100%	0%
C	28	39%	82%	0%
D	28	14%	75%	25%
E	27	37%	40%	30%
F	24	46%	36%	27%
G	24	8%	0%	50%
H	23	43%	80%	10%
I	22	36%	75%	25%
J	19	32%	100%	17%
K	19	53%	90%	10%
L	16	19%	67%	0%
M	15	7%	0%	0%
N	15	40%	0%	0%
O	14	7%	0%	0%
P	10	10%	0%	0%
Q	8	25%	50%	50%
R	7	43%	67%	0%
S	7	29%	50%	50%
T	4	0%	N/A	N/A
U	3	67%	50%	0%
Avg.	18	28%	52%	15%

Note: % OSPD indicates the frequency in which they provided an OSPD diagnosis out of all the evaluations they conducted. % Non-consent indicates, out of total number of OSPD the evaluator provided, how many times non-consent specifier was used. For example, in this dataset Evaluator C provided an OSPD diagnosis in 11 of his or her 28 evaluations. Of those 11 times OSPD was provided, he or she added the specifier 'non-consent,' 9 times (82%). % Hebephilia indicates, out of total number of OSPD the evaluator provided, how many times hebephilia specifier was used. The Non-consent Combined and Hebephilia Combined variables were used for this analysis.

Table 7

Difference in Evaluator Degree and Use of 'Non-consent' Specifier

	Ph.D.	Psy.D.
Never Used (0%)	2	3
Sometimes Used (1 – 61%)	3	2
Often Used (62 – 100%)	8	2
		$X^2 = 2.42$ $p = 0.30$
Never Used	2	3
Used at least Once	11	4
		$X^2 = 1.83$ $p = 0.18$

Table 8

Difference in Evaluator Gender and Use of 'Non-consent' Specifier

	Male	Female
Never Used (0%)	4	2
Sometimes Used (1 – 61%)	2	2
Often Used (62 – 100%)	7	3
		$X^2 = 0.51$ $p = 0.77$
Never Used	4	5
Used at least Once	9	5
		$X^2 = 0.01$ $p = 0.92$

Table 9

Difference in Proportion of Evaluations Conducted and Use of 'Non-consent' Specifier

	Lower Proportion	Higher Proportion
Never Used (0%)	4	1
Sometimes Used (1 – 61%)	3	2
Often Used (62 – 100%)	2	8
		$X^2 = 5.45$ $p = 0.07$
Never Used	4	1
Used at least Once	5	10
		$X^2 = 3.30$ $p = 0.07$

Note: Lower proportion indicates those evaluators who conducted between 3 and 18 evaluations in the current sample. Higher proportion indicates those evaluators who conducted between 19 and 39 evaluations.

Table 10

Difference in Evaluator Degree and use of 'Hebephilia' Specifier

	Ph.D.	Psy.D.
Never Used	8	3
Used at least Once	5	4
		$X^2 = 0.64$ $p = 0.42$

Table 11

Difference in Evaluator Gender and use of 'Hebephilia' Specifier

	Male	Female
Never Used	8	2
Used at least Once	5	6
		$X^2 = 4.10$ $p = 0.04$

Table 12

Difference in Proportion of Evaluations Conducted and Use of 'Hebephilia' Specifier

	Lower Proportion	Higher Proportion
Never Used	7	2
Used at least Once	2	9
		$X^2 = 7.10$ $p = 0.00$

Note: Lower proportion indicates those evaluators who conducted between 3 and 18 evaluations in the current sample. Higher proportion indicates those evaluators who conducted between 19 and 39 evaluations.

Table 13

Use of 'Non-consent' Specifier and Age of Victim in Index Offense

	Non-Consent Specifier	
	No	Yes
10 and Under*	52	0
12	7	2
13 – 14	25	1
15 – 16	17	9
Adult	128	46
Multiple Age Groups	34	6
		$X^2 = 10.04$ $p = 0.04$*

*10 and under age group excluded from chi-square analysis.

Table 14

Use of 'Non-consent' Specifier and Offender Preferred Victim Age

	Non-Consent Specifier	
	No	Yes
10 and Under*	22	0
12	0	2
13 – 14	0	0
15 – 16	2	0
Adult	77	23
Multiple Age Groups	211	43
		$X^2 = 10.75$ $p = 0.01$*

*10 and under age group excluded from chi-square analysis.

Table 15

Use of 'Hebephilia' Specifier and Age of Victim in Index Offense

	Hebephilia Specifier	
	No	Yes
10 and Under	50	2
12	8	2
13 – 14	24	2
15 – 16	24	2
Adult	171	3
Multiple Age Groups	37	3
		$X^2 = 11.35\ p = 0.05$

Table 16

Use of 'Hebephilia' Specifier and Offender Preferred Victim Age

	Hebephilia Specifier	
	No	Yes
10 and Under	21	1
12	2	0
13 – 14	0	0
15 – 16	2	0
Adult	100	0
Multiple Age Groups	239	15
		$X^2 = 6.39\ p = 0.17$

Appendix A

OSPD Labels Provided by Evaluators for Each Case with an OSPD Diagnosis

Table A1

OSPD Labels Provided when both Evaluators *Agreed* on OSPD Diagnosis

Case #	Evaluator 1: OSPD	Evaluator 2: OSPD	Evaluator 1: OSPD Label	Evaluator 2: OSPD Label
1	✓	✓	Non-consent	Non-consent
2	✓	✓	Non-consent	Non-consent
3	✓	✓	Non-consent	Non-consent
4	✓	✓	Non-consent	Non-consent
5	✓	✓	Non-consent	Non-consent
6	✓	✓	Non-consenting Partners	Non-consent
7	✓	✓	Nonconsenting	Nonconsensual Sex
8	✓	✓	Non-consent	[No Label]
9	✓	✓	[No Label]	Non-consent
10	✓	✓	Non-consent	NOS
11	✓	✓	Non-consent	with features of Paraphilic Coercion and Courtship Disorder
12	✓	✓	Biastophilia	Non-consent
13	✓	✓	Biastophilia/Paraphilic Rape	Nonconsenting with Females
14	✓	✓	[No Label]	Biastophilia
15	✓	✓	Non-consenting Partner (NOS)	Non-consent, in a controlled environment
16	✓	✓	Non-consenting Sex	in a Controlled Envrionement
17	✓	✓	Biastophilia	[No Label]
18	✓	✓	Sexting	Hebephilia
19	✓	✓	Non-consent; Hebephilia	Non-consent
20	✓	✓	Hebephilia	Ephebophilia, Hebephilia
21	✓	✓	Adolescent Females	Attraction to Adolescent Females
22	✓	✓	Zoophilia	Bestiality
23	✓	✓	[No Label]	[No Label]

Table A2

Discrepancy when Evaluator One Provided OSPD Diagnosis

Case #	Eval. 1: OSPD	Eval. 2: OSPD	Eval. 1: OSPD Label	Eval. 2: UPD	Eval. 2: Paraphilic D/O	Eval. 2: Sexual Sadism	Eval. 2: Pedophilia
24	✓	X	Non-consent	✓	✓		
25	✓	X	Non-consent	✓	✓		
26	✓	X	Non-consent	✓	✓		
27	✓	X	Non-consent	✓	✓		
28	✓	X	Non-consent	✓	✓		
29	✓	X	Nonconsenting	✓	✓		
30	✓	X	Non-Consenting	✓	✓		
31	✓	X	with Exhibitionistic and probable Pedophilic/Hebephilic Features	✓	✓		
32	✓	X	[No Label]	✓	✓		
33	✓	X	Non-consent	X	✓	✓	
34	✓	X	Non-consent	X	✓	✓	
35	✓	X	Non-consent	X	X		
36	✓	X	Non-consent	X	X		
37	✓	X	Non-consent	X	X		
38	✓	X	Non-consent	X	X		
39	✓	X	Non-consent	X	X		
40	✓	X	Non-consent	X	X		
41	✓	X	Nonconsenting in a Controlled Environment	X	X		
42	✓	X	Nonconsenting Persons	X	X		

43	✓	X	Non-consenting Sex with Females, in remission in a Controlled Environment	X	X	
44	✓	X	features of Non-consenting Persons and Exhibitionism	X	X	
45	✓	X	with Arousal to Non-consensual Sex	X	X	
46	✓	X	Biastophilia	X	X	
47	✓	X	Necrophilia	X	X	
48	✓	X	Nonconsenting Complex: Nonconsent, Force, Violence, Compulsive Use of Pornography and Telephone Scatologia	X	X	
49	✓	X	with features of Hebephilia and Biastophilia	X	X	
50	✓	X	Nonconsensual Sexual Activity with Age Inappropriate Individuals	X	✓	✓
51	✓	X	Sexually Attracted to Young Pubescent Females, in a Controlled Environment	X	✓	✓
52	✓	X	in a Controlled Environment	X	X	
53	✓	X	[No Label]	X	✓	✓
54	✓	X	[No Label]	X	✓	✓
55	✓	X	[No Label]	X	X	
56	✓	X	[No Label]	X	X	
57	✓	X	[No Label]	X	X	
58	✓	X	[No Label]	X	X	
59	✓	X	[No Label]	X	X	

Note: UPD = Unspecified paraphilic disorder.

Table A3

Discrepancy when Evaluator Two Provided OSPD Diagnosis

Case #	Eval. 1: OSPD	Eval. 2: OSPD	Eval. 2: OSPD Label	Eval. 1: UPD	Eval. 1: Paraphilic D/O	Eval. 1: Sexual Sadism	Eval. 1: Pedophilia	Eval. 1: Other
60	X	✓	Non-consent	✓	✓			
61	X	✓	Non-consent	✓	✓			
62	X	✓	Non-consent	✓	✓			
63	X	✓	Non-consent	✓	✓			
64	X	✓	Nonconsent, Sadistic and Exhibitionistic Features	✓	✓			
65	X	✓	Sadistic Features	✓	✓			
66	X	✓	Nonconsenting; in a Controlled Environment	✓	✓			
67	X	✓	with Sadistic Features	✓	✓			
68	X	✓	Hebephilia	✓	✓			
69	X	✓	Hebephilia	✓	✓			
70	X	✓	Non-consent	X	✓	✓		
71	X	✓	Non-consent	X	✓			Voyeurism
72	X	✓	Non-consent	X	✓		✓	
73	X	✓	Non-consent	X	✓		✓	
74	X	✓	Non-consent	X				
75	X	✓	Nonconsent, with Sadistic Features	X	✓	✓		
76	X	✓	Exhibitionistic and Sadistic Features	X	✓	✓		Exhibitionism

77	X	✓	Non-consensual Sexual Activity with Adolescent	X	✓	✓
78	X	✓	Hebephilia	X	✓	✓
79	X	✓	Pedo-hebephilia	X	✓	✓
80	X	✓	Sexually Attracted to Teenagers, in remission in a Controlled Environment	X	✓	✓
81	X	✓	Sexually Attracted to Teenagers, Males and Females, Nonexclusive Type	X	✓	✓
82	X	✓	[No Label]	X	✓	✓
83	X	✓	[No Label]	X		
84	X	✓	[No Label]	X		

Note: UPD = Unspecified paraphilic disorder.

REFERENCES

18 U.S.C. § 4248 (2012).

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

Becker, J. V., Stinson, J., Tromp, S., & Messer, G. (2003). Characteristics of individuals petitioned for civil commitment. *International Journal of Offender Therapy and Comparative Criminology, 47,* 185-195.

Blanchard, R. (2009). Reply to letters regarding pedophilia, hebephilia, and the DSM-V. *Archives of Sexual Behavior*, *38*, 331-334.

Blanchard, R. (2010). The fertility of hebephiles and the adaptationist argument against including hebephilia in DSM-5 [Letter to the Editor]. *Archives of Sexual Behavior*, *39,* 817–818.

Bloom, M., Fischer, J., & Orme, J. G. (1999). *Evaluating practice: Guidelines for the accountable professional* (3rd ed.). Boston: Allyn and Bacon.

Boccaccini, M. T., Murrie, D. C., Caperton, J. D., & Hawes, S. W. (2009). Field validity of the STATIC-99 and MnSOST-R among sex offenders evaluated for civil commitment as sexually violent predators. *Psychology, Public Policy, and Law*, *15*, 278.

Boccaccini, M. T., Turner, D. B., & Murrie, D. C. (2008). Do some evaluators report consistently higher or lower PCL-R scores than others? Findings from a statewide sample of sexually violent predator evaluations. *Psychology, Public Policy, and Law*, *14*, 262.

DeClue, G. (2006). Paraphilia NOS (nonconsenting) and antisocial personality disorder. *The Journal of Psychiatry & Law, 34*, 495-514.

DeClue, G. (2009). Should hebephilia be a mental disorder? A reply to Blanchard et al. (2008). *Archives of Sexual Behavior*, *38*, 317-318.

DeMatteo, D., Murphy, M., Galloway, M., & Krauss, D. A. (2015). A national survey of United States sexually violent person legislation: Policy, procedures and practice. *International Journal of Forensic Mental Health, 14,* 245-266.

Doren, D. M. (Ed.). (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. London: Sage Publications.

Edens, J. F., Boccaccini, M. T., & Johnson, D. W. (2010). Inter-rater reliability of the PCL-R total and factor scores among psychopathic sex offenders: are personality features more prone to disagreement than behavioral features? *Behavioral Sciences & the Law*, *28*, 106-119.

Epperson, D. L., Kaul, J. D., Hout, S., Hesselton, D., & Alexander, W. (2003). *Minnesota Sex Offender Screening Tool-Revised (MnSOST-R) technical paper: Development, validation, and recommended risk level cut scores*. St. Paul: Minnesota Department of Corrections. Available online at http://www.psychology.iastate.edu

Elwood, R. W., Doren, D. M., & Thornton, D. (2010). Diagnostic and risk profiles of men detained under Wisconsin's sexually violent person law. International Journal Of Offender Therapy And Comparative Criminology, 54(2), 187-196.

Frances, A., & First, B. (2011a). Hebephilia is not a mental disorder in DSM-IV-TR and should not become one in DSM-5. *Journal of the American Academy of Psychiatry and Law, 39,* 78-85.

Frances, A., & First, B. (2011b) Paraphilia NOS, nonconsent: Not ready for the courtroom. *Journal of the American Academy of Psychiatry and Law, 39,* 555-561.

Franklin, K. (2009). The public policy implications of "hebephilia" A response to Blanchard et al. 2008 [Letter to the Editor]. *Archives of Sexual Behavior, 39,* 319-320.

Franklin, K. (2010). Hebephilia: Quintessence of diagnostic pretextuality. *Behavioral Sciences and the Law, 28,* 751-768.

Gowensmith, W. N., Sledd, M., & Sessarego, S. (2014). The impact of stringent certification standards on forensic evaluator reliability. 122nd annual meeting of the American Psychological Association, Washington, DC.

Graham, N., & Calkins, C. (2017). *Examining the reliability of DSM-5 diagnoses in SVP evaluations.* Paper presented at the American Psychology and Law Society. Seattle, Washington.

Guarnera, L. A., Murrie, D. C. (in press). Field reliability of competency and sanity opinions: A systematic review and meta-analysis. *Psychological Assessment.*

Guarnera, L. A., Murrie, D. C., & Boccaccini, M. T. (2017). Why do forensic experts disagree? Sources of unreliability and bias in forensic psychology evaluations. *Translational Issues in Psychological Science, 3,* 143-152.

Green, R. (2010). Hebephilia is a mental disorder. *Sexual Offender Treatment*, *5*, 1-7.

Hanson, R. K. & Thornton, D. (2000). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report 99-02). Ottawa, Ontario, Canada: Department of the Solicitor General of Canada.

Hare, R. D. (2003). *Manual for the Hare Psychopathy Checklist-Revised.* (2nd ed.). Toronto, Ontario, Canada: Multi-Health Systems.

Jackson, R. L., & Richards, H. J. (2007). Diagnostic and risk profiles among civilly committed

 sex offenders in Washington state. *International Journal Of Offender Therapy And*

 Comparative Criminology, 51(3), 313-323.

Janus, E. S., & Prentky, R. A. (2003). Forensic use of actuarial risk assessment with sex

 offenders: Accuracy, admissibility, and accountability. *American Criminal Law Review,*

 40, 1-59.

Janus, E. S., & Walbek, N. H. (2000). Sex offender commitments in Minnesota: A descriptive

 study of second generation commitments. *Behavioral Sciences & The Law, 18*(2-3), 343-

 374.

Kansas v. Crane, 534 U.S. 407 (2002).

Kansas v. Hendricks, 117 S. Ct. 2072 (1997).

Kan. Stat. Ann. § 59–29a02 (2011).

King, C. M., Wylie, L. E., Brank, E. M., & Heilbrun, K. (2014). Disputed paraphilia diagnoses

 and legal decision making: A case law survey of paraphilia NOS, nonconsent.

 Psychology, Public Policy, and Law, 20, 294-308.

Knight, R. A. (2010). Is a diagnostic category for paraphilic coercive disorder defensible?

 Archives of Sexual Behavior, *39*, 419-426.

Knighton, J. C., Murrie, D. C., Boccaccini, M. T., & Turner, D. B. (2014). How likely is "likely

 to reoffend" in sex offender civil commitment trials? *Law and Human Behavior*, *38*, 293-

 304.

Levenson, J. (2004). Sexual predator civil commitment: A comparison of selected and released

 offenders. *International Journal of Offender Therapy and Comparative Criminology, 48,*

 638-648.

Lu, Y., Freeman, N. J., & Sandler, J. C. (2015). Predictors of the sex offender civil commitment

      trial outcomes in New York State. Law and Human Behavior, 39(5), 514–524.

Matter of State of New York v. Jason C., 26 N.Y.S.3d 423 (2016).

Matter of State of New York v. Kareem M., 51 Misc.3d 1205 (2016).

Matter of State of New York v. Ralph P., 53 Misc.3d 496 (2016).

McLawsen, J. E., Scalora, M. J., & Darrow, C. (2012). Civilly committed sex offenders: A

      description and interstate comparison of populations. *Psychology, Public Policy, and*

      Law, *18*, 453-476.

Mercado, C. C., Bornstein, B. H., & Schopp, R. F. (2006). Decision-making about volitional

      impairment in sexually violent predators. *Law and Human Behavior*, *30*, 587-602.

McLawsen, J. E., Scalora, M. J., & Darrow, C. (2012). Civilly committed sex offenders: A

      description and interstate comparison of populations. Psychology, Public Policy, and

      Law, 18(3), 453-476. doi:10.1037/a0026116

Miller, A. K., Ruffino, K. A., Boccaccini, M. T., Jackson, R. L., & Murrie, D. C. (2011). On

      individual differences in person perception: Raters' personality traits relate tot heir

      psychopathy checklist-revised scoring tendencies. *Assessment, 18,* 253-260.

Miller, H. A., Amenta, A. E., & Conroy, M. A. (2005). Sexually violent predator evaluations:

      empirical evidence, strategies for professionals, and research directions. *Law and Human*

      Behavior, *29*, 29-54.

Miller, C. S., Kimonis, E. R., Otto, R. K., Kline, S. M., & Wasserman, A. L. (2012). Reliability

      of risk assessment measures used in sexually violent predator proceedings. *Psychological*

      Assessment: Theory, Research, and Treatment, *24*, 944-953.

Money, J. (1999). *The lovemap guidebook: A definitive statement*. New York: Continuum.

Moser, C. (2009). When is an unusual sexual interest a mental disorder? *Archives of Sexual Behavior*, *38,* 323-325.

Mossman, D. (2013). When forensic examiners disagree: Bias, or just inaccuracy? *Psychology, Public Policy, and Law, 19,* 40 – 55.

Murrie, D. C., & Boccaccini, M. T. (2015). Adversarial allegiance among expert witnesses. *Annual Review of Law and Social Science, 11,* 37-55.

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis) agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law*, *15*, 19-53.

Neal, T. M., & Brodsky, S. L. (2016). Forensic psychologists' perception of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law, 22,* 58-76.

Packard, R. L., & Levenson, J. S. (2006). Revisiting the reliability of diagnostic decisions in sex offender civil commitment. *Sexual Offender Treatment, 1,* 1-15.

Perillo, A. D., Spada, A. H., Calkins, C., & Jeglic, E. L. (2014). Examining the scope of questionable diagnostic reliability in Sexually Violent Predator (SVP) evaluations. *International Journal of Law and Psychiatry*, *37*, 190-197.

Plaud, J. J. (2009). Are there "hebephiles" among us? A response to Blanchard et al. (2008). *Archives of Sexual Behavior, 38,* 326-327.

Polaschek, D. L. L. (2003). The classification of sexual offenders. In T. Ward, D. R. Laws, & S. M. Hudson (Eds.), *Sexual deviance: Issues and controversies* (pp. 154 – 171). Thousand Oaks, CA: Sage Publications.

Prentky, R. A., Coward, A. I., & Gabriel, A. M. (2008) Commentary: Muddy diagnostic waters

in the SVP courtroom. *Journal of the American Academy of Psychiatry and Law, 36,*

455-458.

Prentky, R. A., Janus, E., Barbaree, H., Schwartz, B. K., & Kafka, M. P. (2006). Sexually violent

predators in the courtroom: Science on trial. *Psychology, Public Policy, and Law, 12,*

357-393.

Quinsey, V. L. (2010). Coercive paraphilic disorder. *Archives of Sexual Behavior*, *39*, 405-410.

Riddle, D. L., & Stratford, P. W. (1999). Interpreting validity indexes for diagnostic tests: an

illustration using the Berg balance test. *Physical Therapy, 79*, 939-948.

Rind, B. (2013). Critique of Hames and Blanchard (2012), Clancy (2012), and Ryniker (2012) on

hebephilia, anthropological data, and maladaptiveness. *Archives of Sexual Behavior*, *42*,

685-691.

Rind, B., & Yuill, R. (2012). Hebephilia as mental disorder? A historical, cross-cultural,

sociological, cross-species, non-clinical empirical, and evolutionary review. *Archives of*

Sexual Behavior, *41*, 797-829.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and

sample size requirements. *Physical Therapy, 85*, 257-268.

Stephens, S., Seto, M. C., Goodwill, A. M., & Cantor, J. M. (2016). Evidence of construct

validity in the assessment of hebephilia. *Archives of Sexual Behavior, 46*, 301-309.

Stern, P. (2010). Paraphilic coercive disorder in the DSM: The right diagnosis for the right

reasons. *Archives of sexual behavior, 39*, 1443-1447.

Tanner, J. M. (1990). *Foetus into man: Physical growth from conception to maturity.*

Cambridge, MA: Harvard University Press.

The People of the State of Illinois, v. Kevin Stanbridge, 980 N.E.2d 598 (2012).

Thorton, D. (2010). Evidence regarding the need for a diagnostic category for a coercive

       paraphilia. *Archives of Sexual Behavior, 39,* 411-418.

Tromovitch, P. (2009) Manufacturing mental disorder by pathologizing erotic age orientation: A

       comment on Blanchard et al. (2008). *Archives of Sexual Behavior, 38,* 328.

Tucker, D., & Brakel, S. J. (2012). DSM-5 Paraphilic Diagnoses and SVP Law. *Archives of*

       Sexual Behavior, *41*, 533-533.

Wakefield, J. C. (2011). DSM-5 proposed diagnostic criteria for sexual paraphilias: Tensions

       between diagnostic validity and forensic utility. *International Journal of Law and*

       Psychiatry, 34, 195-209.

Zander, T. K. (2007). Commentary: Inventing diagnosis for civil commitment of rapists. *The*

       Journal of the American Academy of Psychiatry and the Law, 36, 459-469.

Zander, T. K. (2008). Inventing diagnosis for civil commitment of rapists. *Journal of the*

       American Academy of Psychiatry and the Law, 36, 459-469.

Zander, T. (2009). Adult sexual attraction to early-stage adolescents: Phallometry doesn't equal

       pathology. *Archives of sexual behavior*, *38,* 329-330.