ADOLESCENT SEXUAL OFFENDING: ASSESSING SEXUAL INTEREST AND EXPLORING TREATMENT TRAJECTORIES AND OUTCOME

by

Julia E. Mackaronis

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology

University of Utah

May 2014

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Julia E. Mackaronis**

has been approved by the following supervisory committee members:

Donald S. Strassberg	, Chair	**10/10/13** Date Approved
Brian Baucom	, Member	**10/10/13** Date Approved
Lisa Diamond	, Member	**10/10/13** Date Approved
David M. Huebner	, Member	Date Approved
Peter M. Byrne	, Member	**10/10/13** Date Approved

and by **Carol Sansone**, Chair of

the Department of **Psychology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

How to best treat individuals who commit child sex offenses is a perennial concern. Many of these perpetrators are adolescents. In order for mental health professionals to adequately address this group's unique treatment needs, further research is necessary. The present research addressed the assessment of sexual interest (which may play an integral role in potential for sexual reoffending) in a sample of adolescents at high risk for sexual reoffending. The project compared results from an unobtrusive method of assessing sexual interest (a viewing time measure) to both a physiologically-based method (a penile plethysmography [PPG] measure), and an actuarially-based assessment of pedophilic interest. Results provided evidence for some overlap between PPG and viewing time assessments, although other evidence suggested that whether or not PPG data are ipsatized may impact the relationship of the data with other assessment modalities. Of particular relevance for clinical assessment, results also suggested that the actuarially-based assessment may have limited utility with adolescents.

There has also been a dramatic increase in treatment programs for adolescents who have committed sex offenses. Despite this increase, treatment dropout rates are often quite high, and more treatment outcome research is needed. To that end, the present study also explored treatment trajectories and outcome, by examining the relationship between pretreatment variables, psychological distress during treatment, and treatment completion, as well as posttreatment outcomes including sexual

reoffending. In a second sample of adolescents in residential treatment for sexual offending, there was no evidence that adolescents who failed to complete treatment, or who had sexual or nonsexual behavior problems following treatment, differed from other adolescents in their trajectories of self-reported distress during treatment. Two trends emerged, however: adolescents who first sexually offended at older ages were slightly less likely to have negative treatment outcomes, whereas adolescents with diagnosed anxiety disorders were more likely to have negative treatment outcomes. Additionally, adolescents with previous academic problems or histories of committing nonsexual crimes experienced different trajectories of self-reported distress than other adolescents, although these trajectories were unrelated to posttreatment outcomes. Implications for research and treatment are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PART I

ASSESSING SEXUAL INTEREST

INTRODUCTION

In the United States, over 90% of individuals who commit child sex offenses are male, and over a third of those males are adolescents when they offend (Finkelhor, Ormrod, & Chaffin, 2009). Research suggests that adolescent sex offenders (ASOs) are distinct from both adolescent general offenders (AGOs; i.e., adolescents with histories of nonsexual criminal offending; Seto & Lalumière, 2010) and adult sex offenders (SOs; Letourneau & Miner, 2005). As a result of both the significant proportion of SOs who are adolescents, and the unique characteristics of ASOs, understanding adolescent sexual offending and treating adolescent sexual offenders is likely critical to reducing rates of child sexual abuse.

Numerous researchers have observed that ASOs are a widely heterogeneous group (e.g., DiCataldo, 2009; Letourneau & Miner, 2005). At the group level, ASOs are not at particularly high risk for sexual recidivism, which is worth noting given that this empirical finding runs contrary to the general public's views about ASOs. In meta-analysis of 63 ASO studies with an average follow-up of nearly 5 years, only 7.08% of the total sample of 11,219 ASOs had been found to have reoffended sexually (Caldwell, 2010). Despite the low overall sexual recidivism rate, a subset of ASOs is nonetheless at much greater risk of continuing to engage in sexually inappropriate behavior across the lifespan. These ASOs are, obviously, of particular concern.

Evidence of this high-risk group has been culled from both retrospective and prospective studies. Across a variety of studies, a subset of adult SOs report first committing sex offenses during adolescence (e.g., Abel, Osborn, & Twigg, 1993; Boutwell & Barnes, 2012; Longo & Groth, 1983). Several retrospective studies of ASOs trace offending-related history back even further. In one sample, adolescents with histories of multiple child sex offenses were more likely to have engaged in inappropriate sexual behaviors as children (Dennison & Leclerc, 2011), and in another sample, those ASOs who had offended sexually both before and after age 12 had also committed more serious sexual offenses (Burton, 2000).

Following such high-risk ASOs through prospective studies yields startlingly different rates of sexual recidivism when compared to ASOs as a whole. Two small samples of particularly high-risk ASOs may be instructive examples. One study was composed of a sample of 14 ASOs in Washington, who, because of their histories of recurrent sexual and often general offending, were subjected to the highest level of community notification (i.e., information regarding their offense history was the most widely disseminated; Lieb, 2006; Schram & Milloy, 1995). At the end of a 3-year follow-up, 43% of the sample had been rearrested for new sexual offenses. Another study was composed of 12 ASOs in Wisconsin who had been recommended for civil commitment, again based on their histories of recurrent sexual offending, often with additional general offending. At the end of a 5-year follow-up, 42% of the sample had been convicted of a new sexual offense, and all sexual reoffenses met penal definitions of serious sexual assault (Hagan, Anderson, Caldwell, & Kemper, 2010). It is worth noting that the most recent study in this area produced different results; comparing two samples of 198 and

144 ASOs who either had or not been recommended for civil commitment, respectively, over a nearly 5-year follow-up period revealed that the two groups sexually offended at similar rates (12% of those recommended for civil commitment versus 17% of those not recommended for civil commitment; Caldwell, 2013). Whether mental health professionals are good at identifying who is high-risk is a separate issue from the fact that a high-risk subgroup of adolescent sexual offenders exists, and obviously warrants considerable attention.

One important area for consideration is how high-risk ASOs might best be recognized and characterized. If a subgroup of ASOs is at unusually high risk for reoffending sexually, what is it about those adolescents that might distinguish them from ASOs more generally? First, it is worth noting that, for ASOs, high-risk for sexual recidivism is quite different from high-risk general recidivism. ASOs as a whole reoffend generally (i.e., reoffend by committing any new, nonsexual crime) at a much higher rate than they reoffend sexually; Caldwell's 2010 meta-analysis found a general recidivism rate of 43.4% across the roughly 5-year follow-up. This high rate of general recidivism that characterizes ASOs as an entire group is likely related to the fact that, like their adolescent general offender (AGO) peers, ASOs often have personality characteristics and life experiences known to be risk factors for general criminal behavior (Seto & Lalumière, 2010; van der Put, van Vugt, Stams, Deković, & van der Laan, 2013).

If all ASOs have a relatively high risk of reoffending generally, what is unique about the much smaller percentage who reoffend sexually? Reoffending may be dependent on opportunity to some degree; one group of researchers recently found an unsurprising link between opportunities to reoffend sexually and sexual reoffenses (Spice

& Viljoen, 2012). There is also evidence, however, that certain ASOs are also at a more internally-based increased risk for sexual recidivism. Specifically, the presence of atypical sexual interest may be particularly important for understanding ASOs at high risk for sexual recidivism. Across studies, one of the most consistent findings about ASOs relates to atypical and problematic patterns of sexual interest (e.g., sexual interest in young children [pedophilia] or arousal to the use of threats and/or force in sex). Seto and Lalumière (2010) observed in their meta-analysis of 50 studies that the presence of atypical sexual interests was the single largest group difference (d = .67) between ASOs and their general offender peers. Atypical sexual interests may play a role in explaining both initial index sex offenses and sexual reoffending. Hanson and Morton-Bourgon (2005), in a meta-analysis of 82 SO recidivism studies, found that atypical sexual interests were the single strongest predictor (d = .30) of sexual recidivism for both ASOs and adult SOs, and McCann and Lussier (2008) found a similar, if somewhat smaller, relationship in a sample of 18 studies of ASO recidivism.

Among ASOs generally, index sexual offenses may or may not be indicative of any broader pattern of sexual interest. The more an offender reoffends sexually, however, the more logical it becomes to assume that the repeated sexual offending is impacted at least in part by sexual offense-specific risk factors, possibly including atypical sexual interest. As Andrews and Bonta (2010) have noted, according to their empirically-supported "need" principle of offender treatment, factors related to offending need to be addressed in treatment. Thus, accurately assessing ASO sexual interest may play an important role in treatment planning and implementation, particularly for ASOs with histories of multiple or particularly serious sex offenses.

Historically, for both ASOs and adult SOs, the most common method of assessing sexual interest has been penile plethysmography (PPG). The assessment procedure entails the man or male teen fitting a measurement device (devices vary from rubber strain gauges to volumetric tubes) around his penis, and his penile blood flow (i.e., sexual arousal) is measured in response to a variety of sexually explicit stimuli. No assessment methodology is perfect, of course, and problems associated with PPG include high rates of men who exhibit no clinically significant arousal to any stimuli ("flat-line" profiles can be due to a variety of factors, including anxiety and intentional suppression of sexual arousal; Kalmus & Beech, 2005; Mahoney & Strassberg, 1991). The percentage of adolescents with clinically uninterpretable responses to PPG assessments has ranged from a low of roughly 1% (Clift, Rajlic, & Gretton, 2009) to above 30% (Becker, Kaplan, & Tenke, 1992). Nonetheless, out of 373 ASO treatment programs throughout the U.S. in one recent survey, over 9% report currently using PPG assessment for ASOs (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010).

In addition to nonresponding, there are two concerns with PPG assessments specific to ASOs. The first concern is ethical. Regardless of age of offender, PPG is a clearly invasive assessment procedure. More specifically for ASOs, some authors (e.g., DiCataldo, 2009; Hunter & Becker, 1994) have pointed out that PPG administrations typically test for atypical sexual interests by presenting not just sexual stimuli, but sexual stimuli including child sexual abuse, sexual coercion, and rape. For the nonnegligible subgroup of ASOs who have experienced sexual, physical, and/or emotional maltreatment (Seto & Lalumière, 2010), PPG assessment might be revictimizing. For ASOs who have no history of maltreatment, exposure to the atypical sexual stimuli that

characterizes PPG assessments may nonetheless be harmful in ways not yet identified. If we also consider the inherently physiologically invasive nature of the assessment, using PPG to assess the sexual interests of ASOs seems less than ideal.

The second concern relates to the validity of what PPG assessments are capturing in ASOs. The utility of PPG for assessing adult SOs is relatively clear; for example, pedophilic sexual interests can be reliably detected even in nonadmitting adult SOs (Blanchard, Klassen, Dickey, Kuban, & Blak, 2001). The few empirical studies of PPG use with ASOs, however, do not yet provide clear guidance as to how ASO PPG results should be interpreted (Becker, Hunter, Goodwin, Kaplan, & Martinez, 1992; Becker, Kaplan et al., 1992; Clift et al., 2009; Gretton, McBride, Hare, O'Shaughnessy, & Kumka, 2001; Hunter, Goodwin, & Becker, 1994; Kaemingk, Koselka, Becker, & Kaplan, 1995; Rice, Harris, Lang, & Chaplin, 2012; Seto, Lalumière, & Blanchard, 2000).

Most recently, Rice and colleagues (2012) found that both ASOs and offense-matched adult SOs displayed greater preferences for child sexual stimuli than individuals in a nonoffending control group. They also observed, however, that ASOs' preferences for such stimuli were weaker than those of the adult SOs. There is also some troubling evidence directly speaking to both the ethical appropriateness and interpretability of PPG assessments for ASOs who have been sexually victimized themselves. In one sample, ASOs who had been sexually abused in childhood were more likely than other ASOs to show indiscriminate (i.e., consistently high) arousal to all stimuli (Becker, Kaplan et al., 1992); there have been similar findings in more recent samples, as well (Murphy, DiLillo, Haynes, & Steere, 2001). Finally, Clift and colleagues (2009) found that, as with adults

(Mahoney & Strassberg, 1991), when ASOs are specifically instructed to suppress their arousal during an assessment, they are effective enough at doing so that the assessment results lose discriminative validity. As a whole, these concerns suggest that PPG assessment and interpretation with ASOs is a complicated endeavor.

An alternative to PPG use in ASO assessment already exists: viewing time assessments. The principle behind the methodology is simple. Basically, men look longer at images of people to whom they are sexually attracted (Israel & Strassberg, 2009; Quinsey, Ketsetzis, Earls, & Karamanoukian, 1996; Rullo, Strassberg, & Israel, 2010). Among adult SOs, viewing time methodologies are effective assessment tools (Harris, Rice, Quinsey, & Chaplin, 1996; Laws & Gress, 2004) with some evidence that viewing time and PPG assessments may yield information of comparable quality and utility (Tong, 2007). In the most recent study of the issue, Mokros and colleagues (2013) found that adult SOs differed from both general offenders and nonoffenders in terms of the amount of time spent looking at child stimuli (in addition to how attractive they rated such stimuli), although classification accuracy was only mediocre. One potential advantage of viewing time assessments is that they may be less susceptible to attempts at willful misrepresentation; it is obvious to men and male teens undergoing PPG assessments that their penile responses are being assessed, whereas it may never occur to someone taking a viewing time assessment that the test is measuring anything other than their subjective ratings of the stimuli.

ASOs frequently undergo viewing time assessments in treatment (over 35% of 373 U.S. treatment programs indicated using such practices; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). Empirical research on viewing time assessment of

ASOs has, however, lagged behind its clinical popularity. Two studies have focused on the Abel Assessment for Interest in Paraphilias (Abel, 1995), with mixed results. An independent team (i.e., not including Abel) found that 2-week test-retest reliability was unacceptably low (Smith & Fischer, 1999), whereas a team led by Abel found that ASOs who looked longer at photographs of children were also more likely to have committed greater numbers of child sex offenses (Abel et al., 2004). Only one study has examined another commonly used viewing time protocol, the Affinity (Version 1.0), and found preliminary evidence for correspondence between viewing time data and offense characteristics (Worling, 2006). No study to date, however, has directly compared ASO PPG and viewing time responses.

For the roughly 9% for ASO treatment programs that still use PPG assessments, switching to viewing time assessments would solve an ethical conundrum, by replacing an invasive and possibly iatrogenic assessment tool with a methodology that is unobtrusive, and therefore less likely to be perceived as unreasonably invasive by the public, as well as ASOs and their caregivers. Viewing time assessments are also both quicker and less expensive. Additionally, if viewing time assessments provide data of potential utility for treatment planning, the majority (over 60%) of ASO treatment programs that use neither PPG nor viewing time assessments may be able to add an easy and clinically useful assessment tool to their treatment approach. To compare the quality of PPG versus viewing time data, as well as to assess the external validity of viewing-time data, we compared PPG and viewing time assessments in a sample of adolescents with histories of sexual offending (typically comprised of multiple offenses). A sample of high-risk ASOs was selected for precisely the reasons discussed above; it is this

population that is of greatest concern regarding ongoing sexual reoffending, and also the same population perhaps most likely to be dealing with atypical sexual interests.

Hypotheses for the present study were as follows: (1) PPG and viewing time data will be significantly positively correlated both with each other and (2) with a measure of pedophilic interest based on offense characteristics (described in further detail in future sections); (3) viewing time data, however, will be more strongly positively correlated with actuarially-assessed pedophilic interest, given that it is an assessment tool less susceptible to willful misrepresentation on the part of those being assessed.

METHODS

Participants and Procedure

Participant data were collected from an existing dataset from a Salt Lake City-based private practice that conducts occasional ASO assessments. (Given the archival and anonymous nature of the dataset, the Institutional Review Board at the University of Utah deemed the project exempt from human subject research reviews.) Participants were 16 males with histories of at least one sex offense against a minor, who received court-mandated psychosexual evaluations between 2004 and 2012 (measures included in the assessment are described in greater detail, below). At the time of evaluation, all participants had already been enrolled in local ASO treatment programs; all evaluations were mandated by the Utah Division of Juvenile Justice Services as a result of the marked severity of the participants' sexual offense history and/or poor progress in treatment. The purpose of the evaluations was specifically to assess participants' level of risk for both sexual and general recidivism.

All participants were between 15 and 20 years old at the time of their psychosexual evaluation (M = 17, SD = 1.55). Participants self-reported victim numbers ranging from 1 to 100 (M = 13.13, SD = 1.41, median = 5.5). A majority of the participants described themselves as heterosexual (n = 10, 62.5%), with 2 identifying as gay (12.5%), 2 identifying as bisexual (12.5%), and 2 stating they were unsure about their sexual orientation (12.5%). Information on their racial/ethnic backgrounds was not

available for analysis. During their assessments, participants completed a brief self-report questionnaire, a MONARCH 21™ Penile Plethysmograph-based assessment, and the Affinity (either Version 2.0 or 2.5) viewing time assessment.

Measures

Self-Report and the Screening Scale for Pedophilic Interest

Prior to PPG assessment, each participant completed a one-page questionnaire of demographic and offense-related questions. Self-report data were coded to score each participant on the Screening Scale for Pedophilic Interest (SSPI; Seto and Lalumière, 2001). The SSPI is a four-item scale that assesses for the presence of pedophilia, yields scores from 0-5 (1 point each for having multiple victims, a victim under the age of 12, or an extrafamilial victim; 2 points for having a male victim) and has been found to be positively correlated with PPG data in ASOs in at least one study (Seto, Murphy, Page, & Ennis, 2003). Utilizing self-reports in forensic contexts always includes the risk of deception; however, three factors about the present study likely mitigate that risk. First, there is evidence that adolescent self-report does not typically diverge wildly from other assessment data; for example, adolescents in one sample self-reported general age preferences (e.g., child versus teen/adult) that corresponded well with the age preferences evident from their viewing time assessments (Worling, 2006). Adolescents in the present sample also knew they were being given a multimethod assessment, and therefore may have been inclined toward honesty if they considered that their self-report might be compared with other measures. Second, complete psychosexual evaluations for this sample included clinician review of related police reports. This means that participants would have had little incentive for distorting offense characteristics that could be easily

cross-checked with third-party information, although it is of course possible that participants failed to reveal additional victims unknown to law enforcement. Third, the type of treatment in which participants engaged is typified by exhaustive discussions of participants' sexual histories, including their sexual offense histories, which means both that treating therapists are another source of third-party information, and that participants would have been practiced in disclosing and discussing the details of their offenses. Adolescents in this sample had relatively high SSPI scores (M = 3.50, SD = 1.51), at least compared with the three other known adolescent samples assessed in one study (Seto et al., 2003).

MONARCH 21™ Penile Plethysmograph

The 90-minute MONARCH 21 PPG assessment (Behavioral Technology, Inc.) utilizes a strain gauge to measures penile circumference change. In the PPG assessment, all participants first viewed a standard 3-minute baseline segment, during which physiological arousal was tracked in the absence of any sexual stimuli. After the baseline segment, all participants viewed 12 standard segments while physiological arousal was tracked via the penile strain gauge: four toddler (ages 3-5) segments, four preteen child (ages 6-11) segments, and four teen (ages 16-18) segments. Each set of segments consisted of two male and two female segments (within genders, one segment described noncoercive sexual contact, whereas the other described coercive sexual contact; we excluded coercive segments from our analyses to avoid conflating possible sources of variance). Each 3-minute segment started with a 90-second audio narrative describing sexual contact, followed by four photographs of partially-clothed individuals of the

corresponding age range and gender. Segment order was randomized during each participant's assessment.

Affinity (Versions 2.0 and 2.5)

The 20-minute Affinity computerized assessment (Pacific Psychological Assessment Corporation) measures both subjective and objective sexual interest.[1] Participants who completed the Affinity 2.0 assessment viewed 56 photographs (28 female and 28 male) of clothed individuals in four age categories: small children (ages 0-5), prejuveniles (ages 6-10), juveniles (ages 11-15), and adults (ages 18 and above). As participants viewed each photograph, they were asked to rate the photographed person's attractiveness on a 15-point gradient with anchors of "Very Unattractive," "Neutral," and "Very Attractive" (yielding subjective self-report scores ranging from -7 to 7 for each photograph), while time spent viewing each photograph was recorded. Participants who completed the Affinity 2.5 assessment viewed 80 photographs (40 female and 40 male) of individuals in the same four age categories used in the Affinity 2.0: small children (ages 0-5), prejuveniles (ages 6-10), juveniles (ages 11-15), and adults (ages 18 and above). As participants viewed each photograph, they were asked to rate the photographed person's attractiveness on a 15-point gradient with anchors of "Unattractive," "Neutral," and "Attractive" (yielding subjective self-report scores ranging from -7 to 7 for each photograph), while time spent viewing each photograph was recorded. For both versions of the Affinity, photograph order was randomized during each participant's assessment. Further, viewing time was measured in two distinct time periods: on-task latency (OTL;

[1] Given that our sample completed testing at a treatment center that only occasionally assesses adolescents, assessments took place over an 8-year span (2004 to 2012), during which time Affinity 2.5 was released. Thus, the use of the two Affinity versions merely reflects the fact that the clinic utilized the most recent versions of assessment instruments available.

time it took a participant to make an attractiveness rating, and which research suggests has more external validity with other assessment measures; Glasgow, 2009) and post-task latency (PTL; time a participant continued looking at a photograph after having made the attractiveness rating).

Data Analysis

Prior to hypothesis testing, we first addressed whether there were adolescents in the sample whose pattern of PPG-assessed responsiveness fell below the acceptable threshold for interpretation. We evaluated the sample using two thresholds for identifying adolescents as PPG nonresponders. The first threshold was if an adolescent's three segments of highest arousal yielded an average circumference change of less than .25 cm; below this threshold, PPG-based diagnostic consistency for pedophilia is no better than chance (Lykins et al., 2010). The second, more conservative threshold was if an adolescent's single segment of highest arousal yielded a circumference change of less than .47 cm (a value which approximates roughly 20% of full erection; Kuban, Barbaree, & Blanchard, 1999).

A final note about the data generated by both PPG and viewing time assessments is necessary here. The research standard for both types of assessment has been to make interindividual comparisons of intraindividually ipsatized, as opposed to raw, data (Glasgow, 2009; Kalmus & Beech, 2005). This standard is by no means non-controversial; although ipsatizing the raw data facilitates interindividual comparisons, the resulting ipsatized scores actually represent relative preferences between the different categories presented in an assessment, as opposed to an individual's absolute preferences (see Fischer & Smith, 1999, for a thorough review of this issue, particularly as it relates

to clinical interpretation of individual assessment profiles). Given the exploratory nature of this first attempt to compare PPG and viewing time data, we conducted all analyses utilizing both raw and individually ipsatized scores from the PPG and viewing-time assessments.

RESULTS

Planned Analyses

The possibility that a portion of adolescents would be classified as nonresponsive to the PPG was addressed during the preliminary data analyses. Regardless of which threshold for nonresponsiveness (described above) was employed, only one participant in our sample (6.3%) evidenced arousal insufficient for interpretation. We conducted subsequent analyses both including and excluding the nonresponding adolescent's data. Only results from analyses including all 16 participants are reported below, as results did not differ between the two approaches. During preliminary data analyses, we also examined the relationship between viewing time variables (see Table 1 for results),[2] which led to our decision to use total task latency (TTL; the total amount of time in seconds an adolescent spent looking at a photo) as the single objective viewing time variable in all subsequent analyses.

We first tested Hypothesis 1, that PPG and viewing time data (both TTL and subjective ratings) would be significantly correlated. The MONARCH 21 and Affinity

[2] Nonparametric statistics were employed because the distributions of both OTL (again, the time in seconds it took a participant to make an attractiveness rating) and PTL (again, the time a participant continued looking at a photograph after having made the attractiveness rating) exhibited marked skewness and kurtosis. We calculated Spearman's rank order correlation coefficients between OTL, PTL, total task latency (TTL; simply the sum of OTL and PTL), and the subjective ratings adolescents made regarding how attractive they found the person in each photograph. All three time-based variables (OTL, PTL, and TTL) were significantly positively correlated with adolescents' subjective ratings (r = .51, .22, and .51, respectively; ps < .001); these findings regarding OTL and PTL match earlier research (Glasgow, 2009). TTL was no more strongly correlated with subjective ratings than was OTL, but the distribution of TTL in our sample had markedly less skewness and kurtosis than either OTL or PTL. For these reasons, TTL was used in all subsequent analyses.

Table 1
Correlations Between Four Viewing Time Variables

	Subjective Attractiveness Rating	OTL	PTL	TTL
Subjective Attractiveness Rating	——			
On-Task Latency (OTL)	.51**	——		
Post-Task Latency (PTL)	.22**	.31**	——	
Total Task Latency (TTL)	.51**	.99**	.53**	——

** $p < .001$

protocols do not include identical stimulus categories, so we limited our comparisons to the most readily comparable categories the two assessment tools share. Six categories are roughly overlapping: two involving male and female small children (ages 3-5 in the PPG assessments, and ages 0-5 in viewing time assessments), two involving male and female preteen children (ages 6-11 in PPG assessments, and ages 6-10 in viewing time assessments), and two involving male and female teens or adults (ages 16-18 in PPG assessments, and ages 18+ in viewing time assessments).[3] We explored both raw and ipsatized versions of the PPG and TTL viewing time data (given that all adolescents made subjective ratings along the same -7 to 7 scale, ipsatized subjective ratings scores were unnecessary).[4] Table 2 summarizes these results. Raw and z-scores from the PPG

[3] For stimuli involving teens, we considered that another possible comparison was between the teen categories (ages 16-18) from the PPG assessment and the young teen categories (ages 11-15) of the viewing time assessment. In preliminary analyses, we computed correlations between the teen PPG categories and both the adult and young teen viewing time categories. A Fischer *r* to *z* transformation indicated that there was no statistically significant difference between the two comparisons. As a result, we compared teen categories (ages 16-18) of the PPG assessment to adult categories (ages 18+) of the viewing time assessment given that these categories might actually overlap to some small degree in a way that comparing 11-year-olds to 18-year-olds would not.

[4] For the PPG data, an adolescent's raw scores were simply the change in penile circumference (measured in cm^2) for each of the six 3-minute segments, whereas their ipsatized scores were *z*-scores for each category based on their mean and standard deviation across the six categories. For the viewing time, each category included multiple individual photographs. An adolescent's raw scores were the means (of both TTL and subjective ratings) of all photographs within each category. Ipsatized viewing time scores were *z*-scores calculated for each category based on their mean and standard deviation in TTL across the four child and two teen categories. The two adult viewing time categories had no direct PPG counterpoints. To make

Table 2
Correlations Between Penile Plethysmography and Viewing Time Variables

	1	2	3	4	5
Penile Plethysmography					
1. Change (cm^2) Raw Scores	——				
2. Change (cm^2) *z*-Scores	.50**	——			
Viewing Time					
3. Total Task Latency (TTL) Raw Scores	.32**	.04	——		
4. TTL *z*-Scores	.09	.07	.66**	——	
5. Subjective Attractiveness Ratings	.05	.03	.69**	.59**	——

** $p < .001$, all other *p*-values > .05.

data were strongly positively correlated with one another ($r = .50$, $p < .001$), obviously, as were raw and *z*-scores from the viewing time TTL data ($r = .66$, $p < .001$). In comparing PPG to viewing time data, however, raw circumferential changes on the PPG and raw TTL were only moderately positively correlated ($r = .32$, $p < .001$), whereas *z*-scores from both measures were not statistically significantly correlated with each other ($r = .07$, $p = .494$). Additionally, adolescents' subjective ratings were strongly positively correlated with both their raw and *z*-scored TTL ($r = .68$ and .59, respectively; *p*s < .001), but those subjective ratings were not correlated with either raw or *z*-scores from the PPG ($r = .05$ and .03, respectively; $p = .647$ and .768, respectively).

We also assessed whether correlations between the PPG and the subjective (ratings) and objective (TTL) viewing time measures varied notably across categories (see Figure 1 for a summary graph; we utilized raw scores here given the lack of

full use of available data, however, we also calculated adolescents' ipsatized viewing time scores as *z*-scores based on their mean and standard deviation in TTL across all eight directly compared categories (male and female small children, preteen children, teen, and adults). This made no significant differences in our correlations between PPG and viewing time data; thus, only analyses utilizing *z*-scores based on the six nonadult categories are reported here.

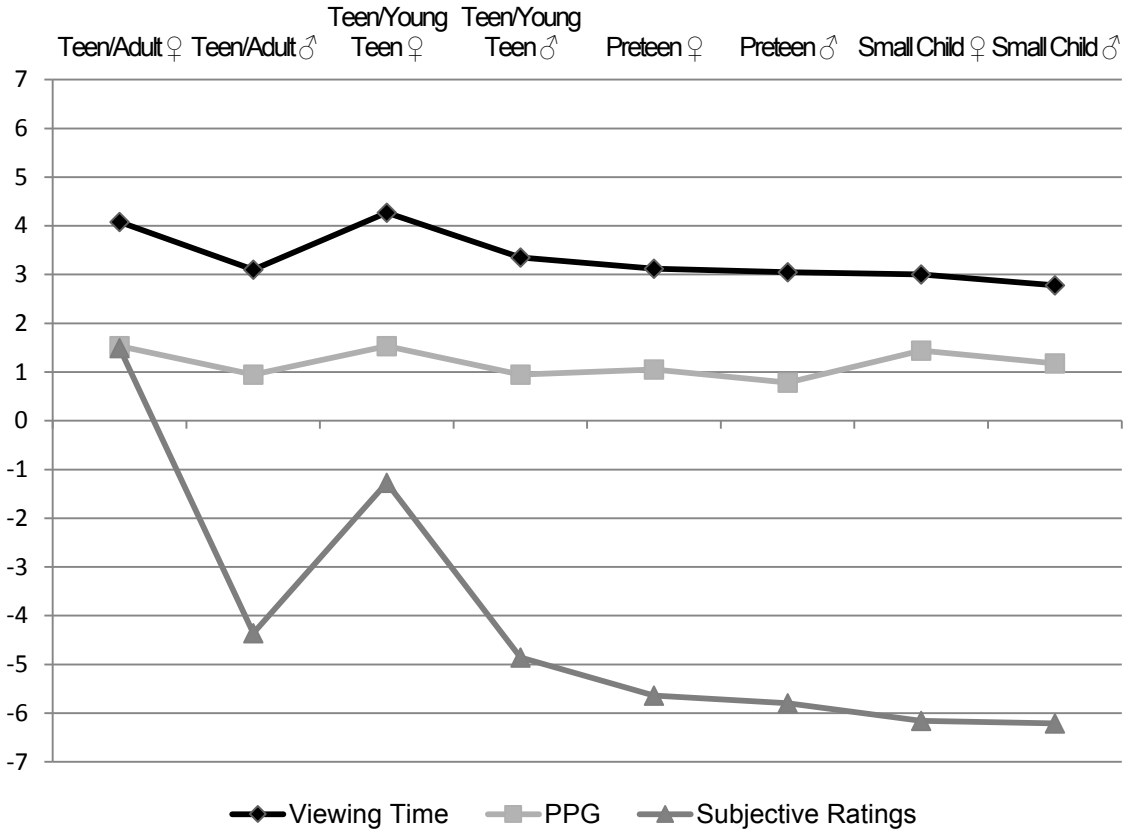

Figure 1. Viewing time total task latency (TTL; measured in seconds), penile plethysmography-based circumferential change (measured in cm^2), and viewing time subjective attractiveness ratings (scale from -7 to 7) across eight categories of sexual stimuli. The lowest possible value for either viewing time or PPG data is 0; all three sources of data are graphed together simply for ease of visual comparison.

correlation between the *z*-scores across measures). Only one of eight comparisons between raw PPG and viewing time TTL data yielded a statistically significant positive correlation (for the male small child category). Five of eight comparisons between raw TTL and subjective ratings yielded statistically significant positive correlations (for the male teen/adult, teen/young teen, preteen, and small child categories, as well as for the female small child categories).

Finally, we calculated Spearman's rank order correlation coefficients to test Hypothesis 2, that PPG and viewing time TTL pedophilic indices would each be significantly positively correlated with SSPI scores in our sample (we examined four possible pedophilic indices per assessment tool).[5] Results are illustrated in Table 3. Unsurprisingly, the four pedophilic indices within each assessment tool were strongly positively correlated with one another (*r*s ranging from .65 to .94, all *p*s < .001); however, none of pedophilic indices were significantly correlated either across PPG and viewing time (*r*s ranging from -.46 to .12, all *p*s > .05) or with SSPI scores (*r*s ranging from -.34 to .03, all *p*s > .05). Given these results, we were unable to test Hypothesis 3, that viewing time data would be more strongly correlated with the SSPI than PPG data.

Post Hoc Exploratory Analyses

It is entirely possible that the lack of significant relationships among PPG, viewing time, and SSPI data in our sample was due to the quite small sample size of 16 adolescents. To further explore this possibility, we took advantage of an available larger

[5] For PPG raw scores, we calculated one pedophilic index as the mean raw score across the four child categories minus the mean raw score across the two teen categories, and another pedophilic index as the highest raw score across the four child categories minus the highest raw score across the two teen categories (for both indices, higher numbers indicate greater interest in younger stimuli). Numerous other studies have utilized the second approach in calculating pedophilic indices; we included the first given the exploratory nature of this comparison between PPG and viewing time data. We repeated these calculations with PPG *z*-scores, TTL raw scores, and TTL *z*-scores.

Table 3

Correlations Between PPG- and Viewing Time-Based Pedophilic Indices and the Screening Scale for Pedophilic Interest (SSPI)

	1	2	3	4	5	6	7	8	9
Penile Plethysmography									
1. Raw Pedophilic Index (means)	——								
2. Raw Pedophilic Index (highest categories)	.92**	——							
3. Ipsatized Pedophilic Index (means)	.94**	.81**	——						
4. Ipsatized Pedophilic Index (highest categories)	.81**	.82**	.83**	——					
Viewing Time									
5. Raw Pedophilic Index (means)	-.10	-.23	-.11	-.20	——				
6. Raw Pedophilic Index (highest categories)	-.28	-.39	-.27	-.46	.87**	——			
7. Ipsatized Pedophilic Index (means)	.12	.00	.06	.01	.87**	.65**	——		
8. Ipsatized Pedophilic Index (highest categories)	-.08	-.16	-.12	-.33	.80**	.89**	.69**	——	
Screening Scale for Pedophilic Interest									
9. SSPI	-.10	.03	-.18	-.08	-.34	-.23	-.28	-.30	——

** $p < .001$, all other p-values > .05.

sample. The MONARCH 21 PPG is used in a variety of locations through the United States, and results from assessments are sent back to the manufacturer of the MONARCH 21 (who also owns the private practice where our smaller sample was assessed). As a result, we had access to PPG data from a national sample of 103 adolescents (including the original 16 participants), as well as self-reported demographic and offense-related characteristics these adolescents provided prior to taking the PPG assessment that also allowed us to generate SSPI scores for this sample.

The 87 adolescents newly included in this larger sample were not significantly different from the original sample in terms of age, as identified by a relevant *t*-test (p = .532). Adolescents added to create the larger sample did have significantly lower SSPI scores (M = 2.23, SD = 1.74) than adolescents in the original, smaller sample (M = 3.50, SD = 1.51; $t(23) = 3.02$, p = .006).[6] This difference in SSPI scores appears to result from the fact that adolescents added to the original sample were less likely to have had multiple victims (47.1% compared to 75.0% in the original sample; $\chi^2(1) = 4.20$, p = .040) and to have had fewer male victims (42.5% compared to 81.2%; $\chi^2(1) = 8.11$, p = .004). This finding, at least in terms of number of victims, provides support for our earlier characterization of the smaller sample as a fairly high-risk group of adolescents. The percentage of adolescents classified as nonresponsive to the PPG using the more stringent .47 cm^2 criteria described above was nearly three times greater in the larger as opposed to smaller sample, although this difference was not statistically significant (21.8% versus

[6] As the degrees of freedom for this *t*-test indicates, Levene's test for equality of variances was statistically significant (F = 4.80, p = .01). The same lack of equivalent variability was also true for a *t*-test assessing group differences in number of victims (F = 6.87, p = .031). Eyeballing the group mean number of victims for the smaller (M = 13.13, SD = 24.11) versus larger sample (M = 5.33, SD = 8.76) suggested a possible group difference that was ultimately nonsignificant ($t(16) = 1.28$, p = .22), perhaps due to the wide variability in victim number in the smaller sample.

6.2%, respectively; $\chi^2(1) = 2.10$, $p = .187$). The smaller sample did appear to have a greater percentage of nonheterosexual adolescents than the larger sample (37.5% versus 11.5%, respectively; $\chi^2(1) = 6.97$, $p = .017$), with the additional 87 adolescents including 1 who identified as gay (1.1%), 7 who identified as bisexual (8.0%), and 2 who stated they were unsure about their sexual orientation (2.3%). The following analyses are from the subsample of 83 adolescents with adequate responsivity to the PPG, although results from the full sample of 103 were virtually identical.

We correlated four PPG-based pedophilic indices[7] with SSPI scores in this larger sample of 83 adolescents (results are displayed in Table 4). The four PPG-based pedophilic indices were, predictably, strongly positively correlated with one another (*r*s ranging from .82 to .92, all *p*s < .001). Contrary to our original hypothesis that PPG data and SSPI scores would be positively correlated, however, both the raw pedophilic index based on means and the ipsatized pedophilic index based on means were weakly, but significantly, *negatively* correlated with SSPI scores ($r = -.26$ and -.25, respectively; both *p*s < .05). This finding is also in the opposite direction of previously published research on the relationship between PPG and SSPI data in adolescents (Seto et al., 2003).

We had visually inspected bar graphs of the full PPG assessment (including additional segments depicting incest, rape, and internet-based sexual activity) for each of the 16 participants in the smaller sample, and observed that several teens had indiscriminately high responding across categories, as has been noted elsewhere in the literature (Becker, Kaplan et al., 1992; Murphy et al., 2001). We thus explored whether

[7] As in our smaller sample, the four pedophilic indices were calculated as (a) the mean raw score across the four child categories minus the mean raw score across the two teen categories, (b) the highest raw score across the four child categories minus the highest raw score across the two teen categories, (c) the mean *z*-score across the four child categories minus the mean *z*-score across the two teen categories, and (d) the highest *z*-score across the four child categories minus the highest *z*-score across the two teen categories.

Table 4
Correlations Between PPG-Based Pedophilic Indices and the Screening Scale for Pedophilic Interest (Larger Sample; N = 83)

	1	2	3	4	5
Penile Plethysmography					
1. Raw Pedophilic Index (means)	——				
2. Raw Pedophilic Index (highest categories)	.87**	——			
3. Ipsatized Pedophilic Index (means)	.92**	.82**	——		
4. Ipsatized Pedophilic Index (highest categories)	.84**	.88**	.90**	——	
Screening Scale for Pedophilic Interest					
5. SSPI	-.26*	-.06	-.25*	-.14	——

*$p < .05$, ** $p < .001$, all other p-values > .05.

the negative correlation between PPG data and SSPI scores might be explained by patterns of indiscriminately high responding. We divided the 83 adolescents into those who scored 0 or 1 on the SSPI (n = 25, 24.3% of the total sample) and those who scored 4 or 5 on the SSPI (n = 43, 41.7%), and then conducted a t-test comparing overall responsivity[8] between those with low versus high SSPI scores. Adolescents who scored a 0 or 1 on the SSPI exhibited lower overall responsivity (M = .61, SD = .52) to PPG assessment than did adolescents who scored a 4 or 5 on the SSPI (M = 1.09, SD = .84; $t(66) = -2.92$, $p = .005$). The negative correlation between PPG-based pedophilic indices and SSPI scores seems less likely to be the result of adolescents who are disinterested in children sexually, and more an issue of greater overall arousability on the part of adolescents who have more extensive and/or serious histories of sexual offending.

[8] We assessed overall responsivity by calculating the mean raw score (cm^2) across the 12 categories to which every adolescent was exposed (including four child and two teen categories depicting coercive sexual activity not incorporated in the primary analyses described above).

DISCUSSION

Our original hypotheses were partially supported in the present study. In this first comparison of PPG and viewing time data in adolescents of which we are aware, PPG and viewing time category-level data were significantly positively correlated with each other, but only when raw as opposed to ipsatized data were utilized. Neither PPG nor viewing time pedophilic indices were correlated with SSPI scores in our smaller, high-risk sample. Unexpectedly, we found two PPG-based pedophilic indices were significantly negatively correlated with SSPI scores in our larger sample. Again in the larger sample, we also found that adolescents with higher SSPI scores evidenced greater overall responsivity to PPG stimuli than did adolescents with lower SSPI scores. Several of our results have implications for the methodology and interpretation of assessing sexual interest in adolescents who have sexually offended, and will be discussed below.

First, readers who are clinicians familiar with using the Affinity in clinical practice may have already noted that our decision to utilize total task latency (TTL) differs from the output data that the Affinity assessment program itself generates and displays in graphs (ipsatized on-task latency or OTL; Glasgow, 2009). Glasgow (2009) has argued that the typically skewed distribution of OTL data is not a concern. Combining OTL and posttask latency (PTL) to examine TTL, however, brought the distribution of viewing time data closer to meeting assumptions of normality, with literally no impact on the strong positive correlation between viewing time and subjective

ratings. Furthermore, research in this area has found evidence that longer viewing times are not simply a result of people taking longer to assign an attractiveness rating to a target who is attractive to them (Imhoff et al., 2010). PTL may reflect a different psychological process than OTL; for example, individuals feeling unsure about how attractive to them an individual slightly outside of their "type" was, might linger over the photo after clicking a rating. Thus, we would consider the methodological question of whether to use OTL, PTL, or TTL to be an incompletely answered question at this time, but one that is certainly worthy of further exploration.

Second is the issue of ipsatization. Had we compared ipsatized data only from the PPG and viewing time assessments, the two tools would have appeared uncorrelated in our smaller sample. This is important to keep in mind, given that ipsatization is the standard approach for both PPG (e.g., Clift et al., 2009) and viewing time (Glasgow, 2009) data. The reasoning behind ipsatization is reasonable; calculating *z*-scores can provide an easily understood metric for comparing results across different people or assessment instruments. If PPG and viewing time data are correlated in terms of raw scores, however, but uncorrelated in terms of ipsatized scores, as they were in our sample, looking only at ipsatized data will provide an incomplete picture of how these two assessment tools are related. Our results suggest that as researchers continue to explore what exactly is the common ground between PPG- and viewing-time based measures of sexual interest, including raw as well as ipsatized data in comparative analyses is warranted.

Our results, particularly those illustrated in Figure 1, add to the body of research showing the value of objective as opposed to subjective measures of sexual interest.

Clinic protocol where these 16 high-risk adolescents were assessed includes explicit instructions to make attractiveness ratings as honestly as possible. The adolescents reported finding the adult women in the Affinity assessment the most attractive by a good margin. Such a strong and singular attraction to adult women may even reflect their actual subjective experience quite well. Overall, however, the more objective PPG and TTL results indicate that these adolescents have the potential to be aroused by a fairly wide variety of stimuli (i.e., the indiscriminately high responding also described by other research teams; Becker et al., 1992; Murphy et al., 2001), despite their typical self-reported narrow preference for adult women. Take a prototypical adolescent from our sample: his subjective self-report is that he is sexually aroused exclusively to adult women, but objective measures indicate he may also be aroused by younger female teens and children. Such an adolescent has two related, but distinct, treatment needs. First, he needs help learning to assess and manage the sexual urges to adult women he acknowledges, but he also needs help learning to assess and manage the sexual urges to female minors he does not yet acknowledge.

Our results also represent the first empirical test of the correlation between PPG and viewing time data, and we found evidence for a moderate correlation (r = .32) between raw PPG scores and raw TTL-based viewing time scores. There are three predictions one might make from this result. First, the correlation we found may be smaller than what would emerge from larger samples; our smaller sample may simply make it difficult to reveal the full extent of the statistical relationship between these two measures. Second, if this moderate correlation is replicated in larger samples, we need to entertain alternate explanations. If PPG and viewing time data are only moderately

positively correlated, what are these two assessment tools measuring, and which measurement is more useful in assessing and treating this population of adolescents? We do know from the current sample that, even with just 16 adolescents, objective and subjective viewing time data are strongly positively correlated (r = .69). Does viewing time data, both objective and subjective, represent "true" sexual interest (which may or may not lead to sexual arousal), whereas PPG data represent a purer measurement of physiological arousability? These adolescents may also experience sexual arousal to so many sexual themes that perhaps PPG or viewing time assessments are insufficiently sensitive to clinically meaningful differences. Whether our results are a product of insufficient sample size, true differences between what PPG and viewing time assessments measure, or artifacts of the instruments themselves, more research on adolescent responses to both types of assessment are strongly suggested.

We found no evidence that PPG or viewing time data are correlated with SSPI scores. This lack of correlation in our sample of 16 adolescents may just reflect the limitations of analyses with such a small group. Sample size was not a concern, however, in our finding of a weak *negative* correlation between SSPI scores and PPG-based pedophilic indices in the larger sample of 83 adolescents, a correlation in the opposite direction of what we expected to find. We have already mentioned one plausible explanation for this finding: adolescents with high SSPI scores showed significantly greater responding across all categories than did adolescents with low SSPI scores. Data from a wide age range support the idea that adolescents as a group exhibit the greatest arousal during assessments (Blanchard & Barbaree, 2005; Kaemingk et al., 1995), but individual differences in arousability may also be important. We already know that

deviant sexual interests are a risk factor for recidivism (Hanson & Morton-Bourgon, 2005), and perhaps the data in our sample reflect that high, nonspecific sexual arousability is meaningfully linked with sexual offending just as is the very specific disorder of pedophilia. For both clinicians and researchers, a cautionary note is that ipsatization can make one adolescent's high, nonspecific arousal look like another adolescent's low, nonspecific arousal, although those two profiles may have very different implications for both assessment and treatment. The adolescent with high non-specific arousal might have truly broad sexual interests or have difficulty disengaging from even not particularly arousing sexual stimuli (e.g., the latter possibility might make it particularly difficult for such an adolescent when faced with a potential trigger for sexual reoffending in the real world). The adolescent with low, nonspecific arousal, on the other hand, may have been attempting to suppress his responses (whether specific or nonspecific) to the PPG assessment.

The categories we used to calculate the PPG-based pedophilic indices may also have contributed to the negative correlation between two of those indices and SSPI scores. Seto and colleagues (2003) found significant positive correlations between PPG data and SSPI scores in three separate samples. Across the three samples, the pedophilic index was calculated as the largest ipsatized response to children minus the largest ipsatized response to adults, individuals over the age of 15, or adolescent peers, respectively. The correlation between a pedophilic index and SSPI scores was strongest (r = .46) in the sample that utilized only responses to adults in the index calculation. It is quite possible that, for adolescents who have offended sexually, arousal to teens can mean a variety of things, on a continuum from normative arousal to peers (or near-peers)

to the very upper age range of truly pedophilic interest. It may simply be more difficult to assess pedophilic interest in adolescents as opposed to in adults, given the smaller age differences between offenders and victims. To that end, our results may well have looked different if the PPG categories available for pedophilic index calculations included adult categories that provide the highest contrast in ages to child stimuli. It appears that, when summarizing adolescent sexual interest, both whether data are ipsatized and what categories of sexual stimuli are ipsatized are important considerations.

This study does have several limitations worthy of consideration. First, particularly with the data from the sample of 16 adolescents, the small sample size had pervasive potential impact on the analyses. All these findings are in need of replication with larger samples, and also with samples incorporating adolescents at varying levels of risk. Second, the ideal comparison for measures of sexual interest would require that all instruments used identical age ranges within categories (e.g., being able to compare segments featuring teens ages 13-17 on both PPG and viewing time assessments); this lack of direct comparison for teen and adult categories was a significant methodological challenge. This comparison was therefore not ideal, and some of the results may be idiosyncrasies resulting from using two assessment tools that were neither designed in tandem nor perfectly analogous. Given that we are unlikely to live in a world of ideal assessments any time soon, more research into what these idiosyncrasies are may nonetheless aid clinicians and researchers using the tools already available to them.

In summary, results from the current study provided evidence for at least some overlap between PPG and viewing time assessments in a small, high-risk sample of adolescents who have sexually offended. The SSPI, on the other hand, may have limited

specificity regarding pedophilic interest in some samples of adolescents. Agreement across assessment measures may also depend heavily on what PPG data are available and how PPG-based pedophilic indices are calculated. Additionally, more research is certainly warranted on adolescents who have offended sexually and also exhibit high, nonspecific patterns of sexual arousal. Assessing sexual interest has important implications for both assessing recidivism risk and treatment planning; our results suggest that there is yet more work to be done to increase the interpretability and better understand the validity of our assessment tools.

PART II

EXPLORING TREATMENT TRAJECTORIES AND OUTCOME

INTRODUCTION

Child sexual abuse continues to be an unfortunately common occurrence in the United States; as a result, treating individuals who have committed child sex offenses remains a critical task for mental health professionals. Sex offender[9] (SO) treatment efforts have largely targeted male perpetrators, in light of the fact that males commit the vast majority of reported child sex offenses (over 90%; U.S. Department of Health & Human Services, 2010). SO treatment efforts have also focused primarily on adults. For example, in the most recent meta-analysis of SO treatment outcome, only 4 out of the 23 studies (roughly 17%) included in the analysis were of adolescent sex offenders (ASOs; Hanson, Bourgon, Helmus, & Hodgson, 2009). This ratio, however, is not especially reflective of actual patterns of child sexual offending; adolescents, in fact, comprise 35.6% of individuals known to have committed child sex offenses (Finkelhor, Ormrod, & Chaffin, 2009).

The existing evidence regarding ASO treatment is encouraging, if preliminary. Across the sample of four studies mentioned above, treatment appeared effective at improving outcomes for ASOs (Borduin, Henggeler, Blaske, & Stein, 1990; Borduin, Schaeffer, & Heiblum, 2009; Cooper, 2000; Worling & Curwen, 2000). Nevertheless, further ASO treatment outcome research is warranted. Two primary reasons for

[9] "Sex offender" can refer to an individual who has offended against either a child or an adult. For the purpose of the present project, the focus is primarily on individuals who have offended against a minor. This term does not connote that such an individual does not also have a history of nonsexual offenses, but merely that he or she is receiving treatment for a child sexual offense.

continuing to expand this body of research exist, and will be discussed in some detail below. First, there is a small subset of ASOs who go on to reoffend sexually; second, although the large majority of ASOs do not reoffend sexually, several factors that distinguish them from adolescent general offenders[10] (AGOs) suggest that ASOs may have important treatment needs above and beyond the fact that they have committed sexual offenses.

For a portion of ASOs, their sexual offending neither starts nor stops at a single isolated incident during adolescence. Adolescents who have committed multiple child sex offenses are more likely than nonrepeat ASOs to have a history of engaging in inappropriate sexual behaviors starting in childhood (Dennison & Leclerc, 2011). There may also be a link between multiple sexual offenses and severity. In one sample, ASOs who self-reported committing sexual offenses both before and after age 12 committed more serious sexual offenses (Burton, 2000). In another sample of ASOs with particularly severe criminal histories, 42% had reoffended sexually by a 6-year follow-up (Hagan, Anderson, Caldwell, & Kemper, 2010). Just as problematic, childhood sexual behavior can precede adolescent sexual offending and adolescent sexual offending can precede adult sexual offending. Among adult SOs, a subset of individuals retrospectively report that they began committing child sex offenses in adolescence, accumulating victims as they aged (e.g., Abel, Osborn, & Twigg, 1993; Longo & Groth, 1983). Given that the sexual recidivism rate among adult SOs is roughly 11% for treated SOs and roughly 19% for untreated SOs (Hanson & Morton-Bourgon, 2009), some individuals with this

[10] The term "general," applied to either adolescent or adult offenders, indicates individuals with criminal convictions for any nonsexual crime, but no convictions for any sexual crimes. Relatedly, ASO recidivism is measured both generally and sexually. An adolescent may reoffend either by committing a nonsexual crime (i.e., general recidivism), or by committing an additional sexual crime (i.e., sexual recidivism).

lifelong pattern of problematic sexual behavior go on to commit additional child sex offenses even after their first such conviction. For this subset of life-course persistent offenders, ASO treatment may present a relatively early opportunity to attempt to divert such individuals from long trajectories of sexual offending, with clear potential benefits both to the offenders themselves, and to their victims.

Authors of the some of the early studies on life-course persistent offenders took something of an alarmist stance toward their results; Longo and Groth (1983) noted in their abstract that "...*many* offenders begin to act out sexually at an early age and if left untreated may escalate to more serious sexual assaults" (p. 150; emphasis added). As has been noted elsewhere (DiCataldo, 2009), however, it would be a mistake to generalize from this subset of adult SOs with lifelong histories of sexually inappropriate behavior to ASOs as a group. This has certainly been borne out by the accumulation of more recent data. Caldwell (2010) culled 63 studies of ASOs, with a total sample of over 11,000 adolescents followed for an average of nearly 5 years, and reported that the weighted mean rate of sexual recidivism was 7.08%. Thus, the proportion of ASOs who might be on a developmental trajectory to become life-course persistent offenders is quite small. This returns us to the second issue raised above: if the large majority of ASOs do not reoffend sexually, why do they need treatment?

ASOs, despite their low base rate of sexual recidivism, need treatment in part because their base rate of general recidivism is quite high. For example, Caldwell (2010) found that the weighted mean rate of general recidivism among his large meta-analytic sample of ASOs was 43.4%; that is, general reoffending was over five times as common as the rate of sexual reoffending he reported. For ASOs, then, treatment may present a

relatively early opportunity to attempt to divert adolescents from trajectories of general offending, with clear potential benefits both to the offenders themselves and to the communities in which they reside.

A related argument is that not only do ASOs need treatment, but that, regardless of their level of risk for either sexual or general recidivism, ASOs deserve treatment. Developmentally-oriented psychologists have long linked detrimental childhood environments with a host of poorer mental and physical health outcomes (see Repetti, Taylor, & Seeman, 2002, for a particularly excellent model and review of the relevant literature). For example, adolescent general offenders (AGOs) are significantly more likely than the average adolescent to have experienced a variety of negative early life experiences (often beginning with and branching out from high-conflict familial environments, which frequently occur in combination with child abuse and/or neglect; Guerra, Williams, Tolan, & Modecki, 2008), and adolescents who offend sexually are even more likely than adolescents who offend generally to have experienced remarkable hardship during childhood! Seto and Lalumière (2010) compared ASOs to AGOs in a meta-analysis of 59 studies; ASOs were more likely to have been sexually, physically, or emotionally abused, neglected, exposed to sexual or nonsexual violence at home, and exposed to sex or pornography.

ASO treatment, then, is important in three ways: (a) reducing sexual recidivism among the small subset of adolescents at high-risk for such behavior, and (b) reducing general recidivism. (c) Recidivism aside, adolescent sexual offending does not exist in a vacuum, and higher frequencies of many negative early childhood experiences may not only contribute to adolescent sexual offending, but also mean that many ASOs have

experienced and continue to be at higher risk for hardships that merit clinical attention and amelioration in their own right.

Given the importance of ASO treatment, the current study focused on explaining treatment outcome among adolescents who have sexually offended. In 2009, there were over 275 community-based and 98 residential ASO treatment programs in the U.S. (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). As noted earlier, there is emerging evidence that treatment for ASOs can be effective (Hanson et al., 2009), but programs are providing ASO treatment faster than empirical research can support such treatment efforts. Research on outcomes for ASOs has diverged somewhat along two separate paths: recidivism versus treatment completion. We will turn first to research on ASO recidivism.

As a point of clarification, different meta-analyses have yielded different estimates of ASO recidivism. Caldwell (2010) reported recidivism rates of 7.08% (sexual) and 43.4% (general), whereas McCann and Lussier (2008) reported rates of 12.2% and 41.7%, respectively (both studies have average follow-up periods of roughly 5 years). Moving forward, a fair generalization would be that, in the 5 years following a first index sex offense, roughly 10% of ASOs reoffend sexually, and upwards of 40% reoffend generally.[11] In addition to replicated statistics regarding rates of recidivism, there is also general agreement on which factors lead to higher risk for sexual recidivism in particular. Across meta-analyses, ASOs with atypical sexual interests (e.g., a primary sexual interest in prepubescent children), prior sex offenses, multiple victims, stranger victims, male victims, antisocial traits, impulsivity, and who are socially isolated, are

[11] Recidivism data of any sort must be conveyed with the caveat that these rates represent only known reoffenses; actual recidivism rates incorporating undetected reoffenses could be significantly higher.

more likely to engage in sexual recidivism than other ASOs (Gerhold, Browne, & Beckett, 2007; McCann & Lussier, 2008; Worling & Långström, 2006). More recent studies continue to provide support for these conclusions (Carpentier & Proulx, 2011; Hendriks & Bigleveld, 2008; Parks & Bard, 2006). For some potential risk factors, current data are mixed. In three samples, ASOs who have been victims of child sexual abuse themselves were at no greater risk for reoffending sexually (Hagan & Cho, 1996; Rasmussen, 1999; Worling & Curwen, 2000), whereas in two other samples, ASOs with such histories were at significantly greater risk for reoffending sexually (Carpentier & Proulx, 2011; Dennison & Leclerc, 2011).

Of particular relevance to the present project, SO treatment completion is also related to risk for sexual recidivism. Among adults SOs, individuals who complete treatment are significantly less likely to reoffend sexually than those who drop out of or are terminated from treatment (Hanson et al., 2002), and the same appears to be true for ASOs (Borduin et al., 1990; Edwards et al., 2005; Rasmussen, 1999; Worling & Curwen, 2000). Treatment dropout among adolescents is a pressing concern, especially given that it appears quite common; in one sample, less than half of ASOs actually completed treatment (Seabloom, Seabloom, Seabloom, Barron, & Hendrickson, 2003).

Which adolescents, then, are at greatest risk for SO treatment dropout? Several studies have provided a preliminary answer to this question. Adolescents who are younger or more impulsive (Kraemer, Salisbury, & Spielman, 1998) or have preexisting emotional and/or behavioral problems (Bremer, 1998) may be less likely to complete treatment. In the most comprehensive exploration to date, Edwards and colleagues (2005) identified 20 risk factors that explained treatment dropout. Some of the factors impacting

dropout in their sample were also those that are linked with higher rates of sexual recidivism (e.g., having male victims, impulsivity). Other factors impacting dropout in their sample included very specific characteristics, such as having an unemployed father and fire setting. Findings such as these raise questions as to whether having an unemployed father or a history of fire setting are uniquely linked to treatment dropout, or are both characteristics simply examples of broader risk factors such as familial instability and antisocial traits, respectively? More research in this area is certainly warranted to identify risk factors that are the most directly related to treatment dropout.

In summary, given the limited number of studies addressing factors that help explain treatment outcome for ASOs, further research is needed to explore the process of treatment, with an eye towards two questions in particular. First, are adolescents who fail to complete treatment distinguishable by their progress in treatment? Second, are adolescents who engage in negative posttreatment behaviors such as nonsexual and sexual reoffending distinguishable by their progress in treatment? This study is designed to address these questions.

Our hypotheses were as follows: (1) adolescents who fail to complete treatment have different trajectories of self-reported distress during treatment than those who do complete treatment (i.e., they will report minimal, if any, reduction in self-reported distress); (2) adolescents who exhibit either nonsexual or sexual conduct problems (including sexual reoffending) at either 6- or 12-months posttreatment will also have different trajectories of self-reported distress during treatment than those who do not engage in problematic posttreatment behavior (i.e., they will report minimal, if any, reduction in self-reported distress); and (3) adolescents with clinician-confirmed

pretreatment psychopathology (e.g., an Axis I diagnosis of a mood, anxiety, or impulse control disorder) will also have different trajectories of self-reported distress during treatment (i.e., they will report minimal, if any, reduction in self-reported distress). Additionally, existing studies of treatment trajectories (described in further detail below) have led researchers to theorize that such trajectories are typically nonlinear. Thus, our final hypothesis is that (4) for adolescents in our sample, the relationship between time in treatment and self-reported distress will be nonlinear, in that adolescents will report the steepest drops in self-reported distress at the start of treatment, with the rate of improvement slowing over time.

METHODS

Participants

Participant data were collected from a Utah-based residential ASO treatment program. (Given the archival and anonymous nature of the dataset, the Institutional Review Board at the University of Utah deemed the project exempt from human subjects research reviews.) Participants were 68 males with histories of at least one child sex offense committed while they were under 18 years of age. The length of treatment varied from 9 to 41 months (in terms of average length of treatment, $M = 17.93$ months, $SD = 6.22$). The treatment program employed a blend of cognitive-behavioral therapy and relapse prevention approaches (the two most common treatment styles with sex offenders; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). All data were compiled by mental health professionals working at the residential program. Please see Table 5 for a summary of relevant sample characteristics, with measures described in further detail below.

Measures

Demographic Characteristics

Clinicians reported collaterally-confirmed information regarding each adolescent's demographic characteristics, including each adolescent's age at the first sexual offense and age at treatment entry, both of which were grand mean centered (Enders & Tofighi, 2007). We incorporated both variables as they were only moderately

Table 5

Pretreatment and Posttreatment Sample Characteristics

Continuous Pretreatment Variables	***M***	***SD***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1. Age at first offense (years)	12.32	2.17	—														
2. Age at entry into treatment (years)	15.01	1.49	.41**	—													
3. SSPI score (range of 0 to 5)	2.84	1.38	-.23	.07	—												
Dichotomous Pretreatment Variables	***n***	**%**															
4. Any mood disorder	33	48.5	-.20	-.09	-.02	—											
5. Any anxiety disorder	10	14.7	-.18	.03	.05	.18	—										
6. Any impulse control disorder	52	76.5	.07	-.14	-.19	-.09	-.06	—									
7. Any disorder involving psychosis	4	5.9	-.06	.09	.00	-.12	.07	-.01	—								
8. Previous placement in foster care	15	22.1	-.16	.02	-.00	.05	.18	.13	.17	—							
9. Previous academic problems	51	75.0	-.01	-.13	-.11	.02	-.05	.08	.00	-.02	—						
10. Previously sexually abused	37	54.4	-.26*	.01	.15	.24*	.13	.05	.10	.35**	.02	—					
11. Previously physically abused	28	41.2	-.02	.07	-.02	.08	.16	.11	.05	.42**	.14	.29*	—				
12. Previous nonsexual crimes	40	58.8	-.03	.15	-.14	.10	.01	.17	.21	.16	.14	.01	.28*	—			
Dichotomous Posttreatment Variables	***n***	**%**															
13. Failure to complete treatment	9	13.2	-.07	.04	-.02	-.03	-.04	.01	-.10	-.10	.03	.10	.03	-.11	—		
14. Nonsexual conduct problems	12	17.6	-.15	.13	-.08	-.02	-.15	.08	-.14	.24	.11	-.06	-.03	.27	-.08	—	
15. Sexual conduct problems	11	16.2	-.02	-.30*	-.08	-.16	-.13	.06	.07	-.06	.19	-.2	.01	.04	.27	-.17	—

Note. All dichotomous variables were coded such that -.05 = absent, and .05 = present. Correlations involving at least one dichotomous variable are Spearman's rank-order correlation coefficients; all others are Pearson product-moment correlation coefficients.

* $p < .05$

** $p < .01$

positively correlated with one another (r = .41, p = .001), and arguments could be made for why either might be more relevant for treatment progress. Additional demographic characteristics included whether the adolescent had a previous placement in foster care, previous academic problems, previous history as a victim of sexual or physical abuse, and previous nonsexual crimes. Each of these qualitative characteristics were represented with dichotomous effect coding as either absent (-.5) or present (.5; Cohen, Cohen, West, & Aiken, 2003).

Screening Scale for Pedophilic Interest

Clinicians also reported police-confirmed offense characteristics for participants' index sexual offense and any additional sexual offenses, including participant characteristics (history of multiple offenses and/or multiple victims) and victim characteristics (age, gender, and relationship of victim to offender). These data were used to score each adolescent on the Screening Scale for Pedophilic Interest (SSPI; Seto & Lalumière, 2001). This four-item scale assesses for the presence of pedophilia and yields scores from 0-5 (1 point each for having multiple victims, a victim under the age of 11, or an extrafamilial victim; 2 points for having a male victim). As with age at first sexual offense, this variable was also grand mean centered.

Pretreatment Psychopathology

Participants' levels of pretreatment psychopathology were assessed clinically during an intake interview conducted by a staff therapist, which resulted in a DSM-IV-TR (APA, 2000) five-axis diagnosis of each participant. Diagnoses were represented with dichotomous effect coding as either absent (-.5) or present (.5) for each of following four variables: (a) presence of any mood disorder diagnosis (e.g., major depressive disorder,

bipolar I or II), (b) presence of any anxiety disorder diagnosis (e.g., generalized anxiety disorder, social phobia), (c) presence of any impulse control disorder diagnosis (e.g., any subtype of attention deficit/hyperactivity disorder, conduct disorder), and (d) presence of any disorder involving psychosis (this category included one adolescent with a diagnosis of thought disorder not otherwise specified, and three adolescents with diagnoses of schizoaffective disorder).

Treatment Progress

During the course of treatment, participants completed the self-report Outcome Questionnaire-45 (OQ-45) at the rate of roughly once every 3 months, with the number of OQ-45s administered varying based on length of treatment. The OQ-45 is a widely-used and well-validated measure of subjective distress (Lambert, Hansen, & Harmon, 2010). The OQ-45 yields a global score and three subscale scores (symptomatic distress, interpersonal problems, and social role dysfunctions); however, a confirmatory factor analysis failed to find support for the distinct properties of the subscales (Mueller, Lambert, & Burlingame, 1998). Only global OQ-45 scores, therefore, were utilized in the present analyses.

Time in Treatment

As noted above, each adolescent completed the OQ-45 roughly quarterly for the duration of treatment, and the date on which any particular OQ-45 score was collected was also recorded. Thus, to examine the longitudinal relationship between treatment progress (as assessed by OQ-45 scores) and time, the date of each adolescent's intial OQ-45 score (which approximates their baseline psychological distress at the start of the treatment program) was coded as 0, and all subsequent time points were calculated as the

number of days between a subsequent OQ-45 administration and the baseline administration. The advantage of representing time in this manner is that it enabled us to look not just at treatment progress at the same discreet time points for each participant (e.g., 3-, 6-, or 12-months into treatment), but treatment progress at a variety of time points (inevitably, some participants filled out the OQ-45 either slightly before or after exactly 3 months treatment).

Additionally, there have been several longitudinal studies of OQ-45 scores during treatment in other populations (Ellsworth, Lambert, & Johnson, 2006; Finch, Lambert, & Schaalje, 2001; Lambert et al., 2002; Vermeersch et al., 2004), and three of the four (Ellsworth et al., 2006; Finch et al., 2001; Vermeersch et al., 2004) have found evidence that treatment trajectories are best fit by quadratic as opposed to linear models (i.e., people experience the sharpest decline in self-reported distress at the beginning of treatment, followed by a more gradual tapering off of symptoms). Thus, our models tested both the impact of number of days in treatment and number of days in treatment squared.

Treament Completion and Outcomes

Treatment completion and outcomes were rated by program clinicians. Given that follow-up data were not collected on adolescents who failed to complete treatment, completion and outcome were considered separately. Treatment completion was treated as a single ordered dichotomous variable for which participants were rated as having: (a) graduated or moved to a lower level of care (coded as -.5) or (b) failed to graduate for one of three reasons (severity of psychopathology, lack of motivation, or moved to a higher level of care; coded as .5). Treatment outcomes were collapsed across the 6- and

12-month clinician-rated follow-up data on 55 of the original 68 participants, and were coded as two ordered dichotomous variables: (a) adolescents who have nonsexual conduct problems (-.5 = absent; .5 = present) and, independently, (b) adolescents who have either sexual conduct problems or an actual sexual reoffense (-.5 = absent; .5 = present).

Data Analysis

The present research was concerned with variance at two levels: change that occurs within an adolescent and change that differs between adolescents. Multilevel modeling (MLM; Raudenbusch & Bryk, 2002) is a style of statistical analysis that was explicitly created to test hypotheses about this kind of multilevel data. MLM enables researchers to explore what proportion of variance occurs at each level of nesting and how variables at each level interact with each other in explaining outcomes of interest.

MLM is also especially useful in studies involving longitudinal change (in the present sample, multiple observations throughout the course of treatment are nested within each individual adolescent) and is considered the most capable methodology for teasing apart within- and between-person variance in such research (Curran & Bauer, 2011). Part of the appeal of MLM is the ease with which complicated datasets can be utilized in their entirety (Raudenbush & Bryk, 2002). Given that the present study is not only multilevel, but longitudinal as well, MLM is particularly well suited to this analytic task (Singer & Willett, 2003), and specific MLM data analytic procedures will be described in further detail. We arrived at the results described below via hierarchical linear modeling (HLM, which is a specific type of MLM, and was employed via HLM 7 for Windows; Raudenbush, Bryk, & Congdon, 2011).

Our models of the data had two levels. The Level 1 model (within adolescent) examined OQ-45 scores longitudinally throughout treatment, and the Level 2 model (between adolescents) examined which pre- and posttreatment adolescent characteristics were related to variance in OQ-45 scores. We first explored the variance in OQ-45 scores by testing the following baseline unconditional means model (this and all subsequent models presented in multilevel model format):

$$\text{Level 1 Model: } OQ_{ij} = \beta_{0j} + r_{ij}$$

$$\text{Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j}$$

where OQ_{ij} represents the OQ-45 score for adolescent j on occasion i, and β_{0j} represents adolescent j's average OQ-45 score. At Level 1, r_{ij} represents within-adolescent variance around each adolescent's average OQ-45 score. At Level 2, γ_{00} represents the average OQ-45 score for the sample, and u_{0j} represents between-adolescents variance around that average OQ-45 score. We next explored whether OQ-45 scores vary significantly over time by testing the following unconditional linear growth model:

$$\text{Level 1 Model: } OQ_{ij} = \beta_{0j} + \beta_{1j}*(DAY_{ij}) + r_{ij}$$

$$\text{Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

where, additionally, DAY represents the number of days adolescent j had been in treatment on occasion i, γ_{10} represents the average change in OQ-45 score per 1 day change in length of treatment, and u_{1j} represents between-adolescents variance around that average change.

As described above, previous research has shown that the decrease in OQ-45 scores is typically quadratic (i.e., it decelerates over time) in nature. Preliminary visual

inspection of Loess-smoothed plots (80% bandwidth, Gaussian kernel) of within-participant treatment trajectories (as recommended by Singer & Willett, 2003) suggested that the relationship between time and OQ-45 scores might be quadratic as opposed to linear in the present sample, as well. Accordingly, we next explored whether OQ-45 scores vary quadratically over time by testing the following unconditional nonlinear growth model:

$$\text{Level 1 Model: } OQ_{ij} = \beta_{0j} + \beta_{1j}*(DAY_{ij}) + \beta_{2j}*(DAYSQ_{ij}) + r_{ij}$$

$$\text{Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

where, additionally, *DAYSQ* represents the number of days adolescent *j* had been in treatment on occasion *i* squared, γ_{20} represents the average acceleration/deceleration of change in OQ-45 scores per one unit change in *DAYSQ*.[12]

Given the large number of possible explanatory variables representing between-adolescents differences, we utilized a graduated analytic strategy similar to the strategy utilized by Atkins and colleagues (2005) in order to add between-adolescents variables to the model. We first considered four classes of Level 2 variables as blocks, exploring their impact on the Level 1 OQ-45 intercept and slope. The four blocks were as follows: pretreatment psychopathology (presence of any mood, anxiety, impulse control disorder, or disorder involving psychosis), potential for reoffending (previous nonsexual crimes, SSPI score, treatment completion/failure), victimization history (previous history as

[12] The careful observer will note that no random effect on *DAYSQ* (e.g., between-adolescents variance around that average acceleration/deceleration of change) is included in the model; given the inherent collinearity between DAY and DAYSQ, the model would not run when a u_{2j} term was incorporated, and it was thus dropped from further analyses.

victim of physical or sexual abuse), and psychosocial variables (age at first offense, age at treatment entry, previous placement in foster care, previous academic problems). The purpose of this block-based analysis is to see if any one type of variable explained more between-adolescents variance than other types.

Second, we utilized the Bayesian Information Criterion (BIC; Raftery, 1995) for model selection within the four blocks, which penalizes models for the number of parameters they include—essentially a lack-of-parsimony penalty (Schwarz, 1978). We tested all possible models within each block (e.g., in a block with the three possible explanatory variables A, B, C, we tested seven total models: A, B, C, AB, AC, CB, ABC), computing BIC for each model.[13] Within each of the four blocks, the variables from the model with the lowest BIC (which also had to be lower than the BIC for the model without Level 2 variables) were then retained for the final model. This approach balances model complexity with explanatory power in variable selection (for a discussion of the strengths of this approach, see Atkins et al., 2005), and allowed us to compare both nested and nonnested models (McCoach & Black, 2008).

The final step in our analyses concerned posttreatment outcomes. As stated in the introduction, we hypothesized that adolescents who exhibited either nonsexual or sexual conduct problems (including sexual reoffending) at either 6- or 12-months posttreatment will have shown different trajectories of self-reported distress during treatment than those who did not engage in problematic posttreatment behavior. Given that follow-up data were available on only a subset of the full sample, we considered posttreatment outcomes

[13] The calculation used for BIC was as follows: BIC = $D + \ln(N)*p$, where D is the deviance statistic for the model in question, the natural logarithm of N refers to the N at Level 2 (the number of adolescents in the sample, equal to either 68 or 55, depending on whether a particular model was exploring the full sample or the subsample of adolescents with follow-up data), and p is the number of parameters (i.e., estimated random effects) in the model.

separately. For this analysis, we tested and compared the BIC for models with all possible combinations of these two posttreatment variables (nonsexual or sexual conduct problems) as compared to the unconditional growth model described above in the smaller sample of 55 adolescents.

RESULTS

Basic HLM Analysis

Results from the baseline unconditional means model indicated that the intraclass correlation coefficient for the present data was equal to .53, meaning that roughly 47% of the variance in OQ-45 scores in this sample is within-adolescent and roughly 53% is between-adolescents.

Results from the unconditional linear growth model (summarized in the "Posttreatment Unconditional Growth Model" column of Table 6) indicated that adding days in treatment to the model explained 15% of the within-adolescent variance. Based on this model, an intercept value for all adolescents was estimated to be an initial score of 57.31 on the OQ-45, $SE = 2.68$, $t(67) = 21.31$, $p < .001$. There was also statistically significant between-adolescents variance in those initial scores, $\chi^2(65) = 384.07$, $p < .001$. Furthermore, the relationship between OQ-45 scores and days in treatment was such that the average OQ-45 score declined by an estimated .02 points for every additional day in treatment, $SE = .01$, $t(67) = -3.43$, $p < .001$. There was also statistically significant between-adolescents variance in that slope, $\chi^2(65) = 114.89$, $p < .001$.

Results from the unconditional nonlinear growth model (that is, adding a quadratic effect of days in treatment) indicated that a nonlinear model is actually a poor fit for the present data. The quadratic term did not statistically significantly impact either the intercept of OQ-45 scores or the slope of OQ-45 scores over days in treatment.

Table 6

Hierarchical Linear Modeling Results for Unconditional Growth and Full Models

	Posttreatment				**Across 6- and 12- Month Follow-Ups**			
	Unconditional Growth Model		Full BIC Model		Unconditional Growth Model		Full BIC Model	
Fixed Effects								
Variable	B (β)	*p*	B (β)	*p*	B (β)	*p*	B (β)	*p*
Intercept (Initial OQ-45 Score)								
Intercept	57.31	<.001	58.90	<.001	57.20	<.001	57.15	<.001
Anxiety			8.24 (.11)	.27				
Thought			7.84 (.08)	.54				
Crimes			-7.20 (-.14)	.16				
Tx Failure			-.33 (.00)	.94				
Foster Care			-4.91 (-.08)	.49				
Acad			15.75 (.26)	.01				
Conduct							-5.32 (.09)	.52
Sexual							4.82 (.08)	.62
Slope (OQ-45 Scores Across Days in Treatment)								
Intercept	-.02 (-.17)	<.001	-.03 (-.25)	.06	-.02 (-.17)	.001	-.03 (-.25)	.04
Anxiety			.01 (.05)	.69				
Thought			-.01 (-.05)	.49				
Crimes			.02 (.13)	.08				
Tx Failure			.01 (.06)	.52				
Foster Care			-.02 (-.12)	.21				
Acad			-.01 (-.06)	.21				
Conduct							-.01 (-.06)	.69
Sexual							-.01 (-.06)	.89
Random Effects								
	Variance Component	*p*	Variance Component	*p*	Variance Component	*p*	Variance Component	*p*
r_{ij}	248.41		247.36		245.58		246.41	
u_{0j}	384.07	<.001	361.15	<.001	384.83	<.001	393.02	<.001
u_{1j}	.03	<.001	.001	<.001	.001	<.001	.001	<.001
Deviance Statistic (-2*LL*)								
	3765.03		3752.21		3047.54		3047.01	
Bayesian Information Criterion (BIC)								
	3781.90		3769.08		3063.57		3063.04	

Note. B = estimated (nonstandardized) fixed-effect regression coefficient; β = standarized regression coefficient (calculated as $\beta = B*(SD_{Variable}/SD_{OQ\text{-}45})$; Raudenbush, Rowan, & Kang, 1991); Anxiety = any anxiety disorder; Thought = any disorder involving psychosis; Crimes = previous nonsexual crimes; Tx Failure = failure to complete treatment; Foster Care = previous placement in foster care; Acad = previous academic problems; Conduct = nonsexual conduct problems during follow-up; Sexual = sexual conduct problems during follow-up; r_{ij} = within-adolescent variance around each adolescent's average OQ-45 score; u_{0j} = between-adolescents variance around the average OQ-45 score at the beginning of treatment; u_{1j} = between-adolescents variance around the average change in OQ-45 score per 1 day change in length of treatment.

Additionally, in comparing the deviance statistic (a measure of model fit, with lower numbers indicating a better fit)[14] between the two models, the nonlinear model was a worse fit that the linear model. The same was also true when comparing the BIC of each model; the nonlinear model yielded a very strong deterioration in overall fit as compared to the linear model (ΔBIC = 16.31, with increases in BIC indicating decrements in fit).[15] There are several possible reasons for the lack of a quadratic effect in the present sample. First, the sample size might simply have been too small to detect such an effect. Second, the research that supported the presence of a quadratic effect of time utilized OQ-45 scores that were collected on a once-per-session (roughly weekly) basis, as opposed to the quarterly OQ-45 assessments in the present sample. Given that nonlinear models have been characterized by a steeper drop-off in OQ-45 scores at the beginning of treatment, the time scale of the observations in this sample may simply have been too long to detect rapid changes early in treatment. Thus, although there is evidence from previous research for a quadratic effect of time, we did not retain the number of days in treatment squared (represented as *DAYSQ* above) in subsequent models.

Blocks of Variables: Pretreatment Psychopathology, Potential for Reoffending, Victimization History, and Psychosocial Variables

Once we established that the unconditional linear growth model would serve as our new baseline comparison model for evaluating the utility of additional variables, we tested the collective contributions of our four classes of between-adolescents variables (pretreatment psychopathology, potential for reoffending, victimization history, and

[14] For any model, deviance = -2*(log-likelihood of the current model – log-likelihood of the saturated model), sometimes abbreviated as -2*LL*, meaning the deviance statistic is a comparison of how much worse the current model is than a model fitting the data perfectly (Singer & Willett, 2003, p. 117).
[15] In evaluating differences in the BIC between these two models, we utilized the following guidelines: differences of 0-2 as weak, 6-10 as strong, and over 10 as very strong (Raftery, 1995).

psychosocial variables). As noted above, Table 5 provides a summary of these variables, and age at first offense, age at treatment entry, and SSPI scores were centered around the grand mean prior to analyses. In terms of BIC for each of the four blocks we tested, all four blocks resulted in more poorly fitting models than the unconditional linear growth model.[16] This finding highlights the advantage of conducting analyses in such a way that variables are examined both collectively, as was done in this block-based analysis, and separately, as was done next.

Posttreatment Model Selection With BIC

The full model was constructed with the BIC variable selection procedure described above. The results from this full model are presented in Table 6 under the "Posttreatment – Full BIC Model" column; out of 13 possible additional variables, 6 were included (presence of any anxiety disorder, presence of any disorder involving psychosis, previous nonsexual crimes, treatment completion/failure, previous placement in foster care, and previous academic problems). Readers will note that relatively few included variables had statistically significant relationships with the OQ-45 intercept and slope. The decision to include nonsignificant variables was based on our BIC comparisons. As compared to the unconditional linear growth model, the full model resulted in a very strong improvement in fit (ΔBIC = -12.82). This improvement indicates that, at least in this sample, it was worth including nonsignificant variables that nonetheless contributed to model fit.

Two interactions are worthy of note. First, adolescents with previous academic problems had initial OQ-45 scores nearly 16 points higher than adolescents without such

[16] As Atkins and colleagues (2005) have noted, a model with additional variables that explains less overall variance is a poorly fitting model indeed!

problems, *SE* = 6.20, *t*(61) = 2.54, *p* = .014 (see Figure 2). Second, there was a non-significant trend such that adolescents who had previously committed nonsexual crimes had a less steep decrease in OQ-45 scores over time, B = .02, *SE* = .01, *t*(61) = 1.77, *p* = .083 (see Figure 3). The lack of two interactions is also of note; contrary to our hypotheses, failure to complete treatment was not significantly related to either initial OQ-45 scores, or change in OQ-45 scores over time. All other within-adolescent and between-adolescents effects were nonsignificant.

Follow-Up Model Selection With BIC

This full model on the subset of 55 adolescents with 6- and 12-month follow-up data was also constructed with the BIC variable selection procedure described above. The results from this model are presented in Table 6 under the "Across 6- and 12-Month Follow-Ups – Full BIC Model" column. Out of the two possible additional follow-up variables (nonsexual or sexual conduct problems), the model including both had the lowest BIC. As can be seen in Table 6, however, this full model was only a very weak improvement in fit as compared to the unconditional linear growth model (ΔBIC = -.53). Unlike in our full model for the full sample, neither posttreament follow-up variable significantly contributed to our understanding of changes in OQ-45 scores over the course of treatment.

Post Hoc Exploratory Analyses

The primary disadvantage of the mulilevel modeling described above is that we utilized OQ-45 scores as our outcome, exploring if posttreatment outcomes such as reoffending explained treatment trajectories. What happens, then, if we use posttreatment outcomes as outcomes? We examined this question through exploratory logistic

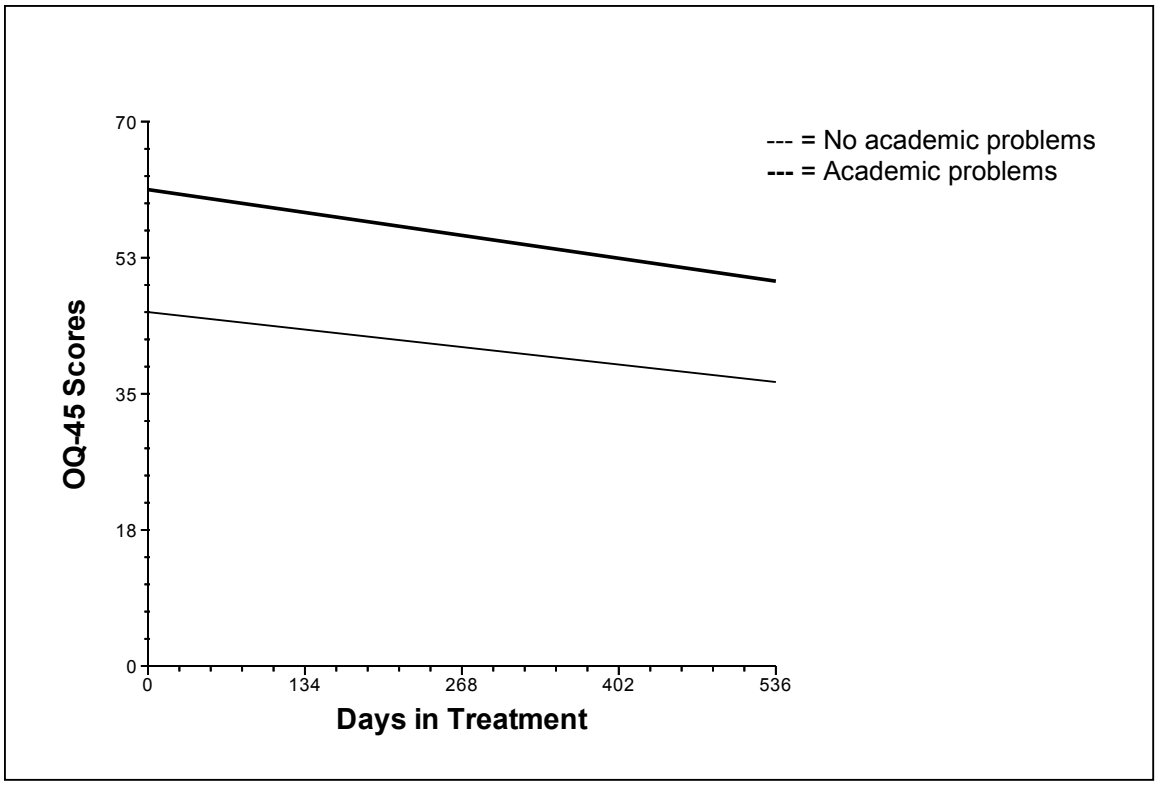

Figure 2. OQ-45 scores over time for adolescents with and without previous academic problems.

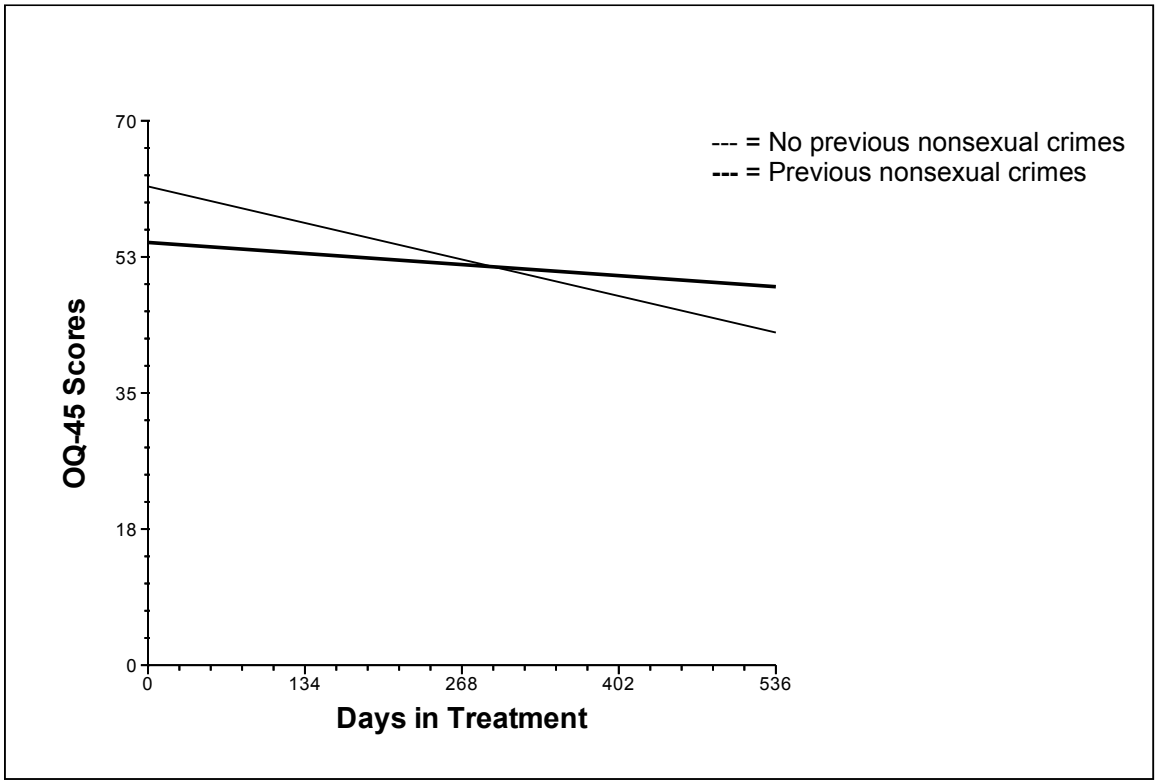

Figure 3. OQ-45 scores over time for adolescents with and without histories of committing previous nonsexual crimes.

regression. We chose to consider all pretreatment measures as possible explanatory variables, regardless of whether they were included in the final multilevel model, given that a pretreatment variable might have a nonsignficant relationship with OQ-45 scores, but a significant relationship with one of our postreatment variables. We also wanted to assess the possibility that either initial or final OQ-45 scores might explain treatment outcomes.

We thus examined the following 12 variables as possible explanatory variables: the presence of any mood, anxiety, impulse control disorder, or disorder involving psychosis; previous nonsexual crimes; SSPI score; previous history as victim of physical or sexual abuse; age at first offense; age at treatment entry; previous placement in foster care; previous academic problems; and both the initial (M = 54.34, SD = 24.09) and final (M = 47.34, SD = 23.17) OQ-45 scores (these scores and all other continuous variables were grand-mean centered prior to analysis; dichotomous variables were re-coded as 0 for absent and 1 for present). Specific negative treatment outcomes were relatively rare in our sample, so we collapsed the three negative treatment outcomes (failure to graduate, nonsexual conduct problems, or sexual conduct problems) into one "any negative treatment outcome" category. In our sample, 29 adolescents had negative treatment outcomes, whereas 39 did not.

We conducted a stepwise logistic regression (forward selection, with removal based on likelihood-ratio statistics) to predict any negative treatment outcome with the 12 variables described above. Two variables, age at first offense and presence of an anxiety disorder, were retained in the final model. An omnibus test indicated that the final model was a significant improvement over a model with no additional variables, $\chi^2(2)= 6.71$, p =

.035, and a Hosmer-Lemeshow test indicated that the final model was a good fit for the data, $\chi^2(6)= 5.90$, $p = .435$.[17] Despite the final model being an improvement in fit, both age at first offense and the presence of an anxiety disorder only trended towards statistical significance in their ability to explain negative treatment outcomes, and should thus be interpreted cautiously. In the case of age, for every 1 month increase in an adolescent's age at first offense, he was .78 times less likely to have a negative treatment outcome ($\chi^2 = 3.72$, $p = .054$, 95% CI = .61 to 1.00). In the case of anxiety, adolescents with any diagnosed anxiety disorder were 4.83 times more likely to have a negative treatment outcome ($\chi^2 = 3.26$, $p = .071$, 95% CI = .873 to 26.70).

[17] Unlike many statistical tests, a nonsignificant *p*-value on the Hosmer-Lemeshow test is indicative of good fit.

DISCUSSION

In the broadest strokes, our final model provides some useful information regarding how adolescents in treatment for sexual offending fare over the course of that treatment. The adolescents in our sample entered treatment reporting considerable subjective distress, and that distress lessened in a gradual, linear fashion as treatment progressed. This is not an unimportant finding. As mentioned earlier, the creation of new treatment programs for adolescent sexual offending has far outpaced treatment outcome research. Thus, the finding that this program did seem to help decrease adolescents' subjective distress over time is heartening. As clinicians, we hope that all individuals benefit from psychotherapy, but as an ethical matter, it seems particularly important that individuals (especially minors) who enter mandatory treatment might gain some benefit from the requirement above and beyond reducing their risk for recidivism.

Furthermore, using our multilevel model to compare within- versus between-adolescent differences, we found that just over half of the variance in OQ-45 scores occurred between-adolescents. As in many other domains of psychology, individual differences matter. In our sample specifically, several individual differences (discussed in more detail below) exerted an impact on adolescents' subjective distress either at the start of treatment or in the rate of change in subjective distress over time.

Our analyses, however, did not find evidence for any of our four original hypotheses. Considering our first two hypotheses, we found that (a) adolescents in our

sample who failed to complete treatment did not have significantly different trajectories of subjective distress during treatment, and (b) adolescents in our sample who exhibited either nonsexual or sexual conduct problems at follow-up also did not have significantly different trajectories of subjective distress during treatment. These failures to reject the null hypotheses may be specific to our relatively small sample, but are nonetheless somewhat disconcerting. The primary purpose of treating adolescents who have sexually offended is to reduce the chance they will reoffend sexually. Two secondary purposes are to reduce the chances they will fail to complete treatment (particularly given that treatment failure is related to sexual recidivism; Borduin et al., 1990; Edwards et al., 2005; Rasmussen, 1999; Worling & Curwen, 2000) or will commit future nonsexual criminal offenses. In many situations, a therapy client's subjective distress can be used as a barometer of therapy effectiveness (e.g., an anxious individual's Beck Anxiety Inventory scores should decrease over time if therapy targeting that anxiety is working). It appears, at least in our sample, that an adolescent's subjective distress did not serve a similar barometric function for posttreatment outcomes.

Our failure to find a relationship between OQ-45 scores and negative posttreatment outcomes raises an etiological question: is subjective distress related to treatment failure, or sexual or general reoffending? In the case of sexual reoffending, our primary treatment target, existing research has addressed this question only in parts. On the one hand, Seto and Lalumière (2010) found that adolescents who have sexually offended have higher rates of depression, anxiety, and social isolation than do their general offender peers.[18] On the other hand, in a meta-analytic sample comprised mostly

[18] This also raises questions about causality and directionality; did these adolescents become depressed, anxious, and socially isolated before, during, or after their sexual offense(s)?

of adult sex offenders, loneliness (which is plausibly symptomatic, although not diagnostic, of depression, anxiety, or social isolation) was virtually unrelated to sexual recidivism (d = .03; Hanson & Morton-Bourgon, 2005). Given that studies of sexual recidivism in adolescents have typically not included variables measuring subjective distress, it is thus currently impossible to answer more complicated questions. For example, suppose that subjective distress is *in*directly related to sexual reoffending through a construct like self-regulation (i.e., subjectively distressed adolescents have a compromised capacity for self-regulation, and therefore for resisting opportunities to reoffend sexually). Such an explanation may not be implausible. In their meta-analysis, Hanson and Morton-Bourgon (2005) did find that self-regulation problems were moderately related to sexual recidivism (d = .37), and that finding dovetails nicely with recent empirical support for self-(dys)regulatory pathways to sexual offending in adults (Kingston, Yates, & Olver, 2013). It is possible, of course, that our findings simply indicate that subjective distress has no significant impact on sexual reoffending for adolescents; it is also possible, however, that an indirect relationship does exist, but is more difficult to measure than a direct relationship would be.

If it is indeed true that subjective distress is not meaningfully linked, directly or otherwise, to negative treatment outcomes in adolescents who have sexually offended, we need research to shed light on what assessment tools (either pretreatment or during treatment) would enable clinicians to anticipate negative posttreatment outcomes in this population. One possibility is simply that the OQ-45, which was designed to be a broad measure of self-reported subjective distress, does capture that distress (and its typical decline over the course of treatment), but fails to capture other psychological variables

that are actually predictive of treatment failure and general or sexual reoffending. Measures of sexual recidivism risk for adolescent sex offenders certainly do exist (for a meta-analysis of research on the four most widely used measures, see Viljoen, Mordell, & Beneteau, 2012), but these measures often rely on *static* (i.e., unchanging) risk factors. In an ideal world, treatment for adolescent sexual offending would focus on and impact *dynamic* (i.e., potentially changeable) risk factors. A potentially fruitful area for further research would be to conduct multilevel modeling similar to the analyses presented here, except using a risk assessment tool that taps dynamic risk factors for sexual offending as the outcome measure. Of the four currently available for use with adolescents, the Estimate of Risk of Adolescent Sexual Offense Recidivism (ERASOR; Worling & Curwen, 2001) is arguably the best suited for the task, given that four out of its five subscales focus specifically on dynamic risk factors. Examining ERASOR scores over the course of treatment might also shed light on risk for general reoffending (Viljoen et al., 2003), but it is less clear whether the measure would have any meaningful relationship with an adolescent's risk for treatment failure. Future research aside, the present results indicate that how we might assess the risk of negative posttreatment outcomes, let alone intervene when such risk is present, is a largely unanswered question.

Considering our third hypothesis, we failed to find that adolescents in our sample with diagnosed and/or noteworthy pretreatment psychopathology showed significantly different trajectories of subjective distress during treatment than did other adolescents. This finding may be of particular use to clinicians, as it suggests that treatment providers ought not to be alarmed by the often quite high rates of psychopathology in this population. Adolescents, regardless of presenting psychopathology, experienced

comparable reductions in subjective distress over the course of treatment. Again, however, we are left with the question of what exactly *is* meaningfully related to posttreatment outcomes.

Our post hoc logistic regression more directly assessed possible relationships between pretreatment variables and posttreatment outcomes, and provided evidence for two trends: (a) adolescents who are older when they first commit a sexual offense may be less likely to have negative posttreatment outcomes, and (b) adolescents with diagnosed anxiety disorders may be more likely to have negative posttreatment outcomes. These trends make some intuitive sense. Adolescents who first offend at younger ages may have more serious predilections for offending that will therefore be more difficult to treat, and this finding replicates results from several earlier studies (Kraemer et al., 1998; Miner, 2002). Anxiety disorders could certainly interfere with an adolescent's ability to engage in treatment, particularly if addressing chronically high anxiety is not a focus of treatment itself. This second finding, however, brings us back to the question of the relationship between psychopathology, subjective distress, and negative posttreatment outcomes. Our multilevel model specifically addressed the possibility that adolescents with anxiety disorders would experience different treatment trajectories (e.g., more distress at treatment entry; more modest reductions over time), and we found no evidence of such differences. If anxiety does influence treatment outcomes negatively, but not in a manner related to subjective distress, this points to the possibility that psychopathology may impact adolescents' propensities for negative posttreatment outcomes in indirect (and not readily apparent) ways.

Regarding both trends that emerged from the logistic regression, our sample is smaller than what is typically used in this form of analysis, and the findings described above were nonsignificant trends; these trends must be therefore be considered as tentative, at best, requiring replication before being taken too seriously. The 10 other variables (presence of any mood, impulse control disorder, or disorder involving psychosis, previous nonsexual crimes, SSPI score, previous history as victim of physical or sexual abuse, age at first offense, age at treatment entry, previous placement in foster care, previous academic problems) showed no ability to explain negative posttreatment outcomes. In our sample, at least, differentiating those who did well after treatment from those who did not was a rather elusive goal. Of course, achieving this goal was significantly hampered by the modest number of adolescents available, particularly those evidencing negative outcomes.

Finally, regarding our fourth hypothesis, the relationship between time in treatment and self-reported distress was linear as opposed to the predicted nonlinear associations, for all the adolescents in our sample. As described above, when nonlinear models provided the best fit for changes in OQ-45 scores over time in previous studies, scores were typically measured at much shorter intervals than in our sample (e.g., once per week; Finch et al., 2001). Thus, one explanation for the linear change in OQ-45 scores over time evident in our sample may be that the time between OQ-45 administrations was too long to detect steeper, nonlinear decreases occuring at the beginning of treatment. Another explanation might be that frequent OQ-45 measurements reveal a steeper and short-lived initial trajectory that has relatively little long-term clinical relevance.

The analyses did reveal two additional relationships unrelated to any specific hypotheses. First, adolescents with previous academic problems reported significantly more subjective distress when entering treatment. This initial distress did not impact their treatment trajectories over time; as Figure 2 illustrates, however, this means that they also left treatment reporting more subjective distress. This finding suggests that it may be particularly important to consider academic ability in treatment planning and implementation; even simple written therapy homework assignments might prove daunting for an adolescent with a history of doing poorly in school. Previous academic problems might also be an indirect measure of motivation. If an adolescent has done poorly in school simply due to lack of effort, he might certainly be similarly uninspired to apply himself to treatment. Second, adolescents who had previously committed nonsexual crimes exhibited a trend towards experiencing a less steep decrease in subjective distress over time. Adolescents with previous criminal histories might be less invested and involved in their own treatment. Another possibility is that adolescents who had committed previous nonsexual crimes were more likely to also have antisocial traits, which are notoriously difficult to address in treatment (although one might also expect that adolescents higher in antisociality might report less distress overall). With both of these patterns, however, any explanation we might offer would make a great deal more sense if previous academic problems and nonsexual crimes were related not only to differences in OQ-45 scores, but also to negative posttreatment outcomes. We found no evidence of such relationships in our sample. We return again to the question of the relationship among pretreatment variables, subjective distress, and negative treatment outcomes. Perhaps these odd results indicate that previous academic troubles and

nonsexual crimes impact an adolescent's ability to benefit *psychologically* from treatment (as evidenced by differences in beginning OQ-45 scores and/or the trajectory of OQ-45 scores over time), while simultaneously exerting little to no impact on an adolescent's ability to benefit *criminologically* from treatment (as evidenced by the lack of relationship between these two variables and treatment failure, and general or sexual reoffending). This is an intriguing area for future research; are there treatment targets that would benefit these adolescents both psychologically and criminologically?

This study did have several noteworthy limitations. First, the sample size was relatively small, particularly given the sophisticated analytic strategies we sometimes employed.[19] Our failure to find a relationship among pretreatment factors, treatment trajectories of subjective distress, and posttreatment outcomes may simply reflect the fact that the sample size was too small to detect such relationships. Second, as a sort of mixed blessing, negative posttreatment outcomes were rare. Adolescents in our sample failed to complete treatment and had nonsexual conduct problems at rates that were absolutely low (12.1% and 18.2%, respectively), as well as relatively low in comparison to rates reported in other samples of ASOs (Caldwell, 2010; Eastman, 2005; Edwards et al., 2005; Hendriks & Bigleveld, 2008). Negative posttreatment outcomes may have been relatively low in our sample because of the relatively short follow-up period; for example, the higher recidivism rates that Caldwell (2010) found in his meta-analysis were based on an average follow-up of nearly 5 years. Adolescents in the present study did have sexual conduct problems during the 1-year follow-up period at a rate (16.7%) that is even slightly higher than what is found in the literature for both ASOs and adult SOs

[19] Given the long-term nature of the treatment program, it took several years to collect even the small sample analyzed here.

(Caldwell, 2010; Hanson & Morton-Bourgon, 2005, 2009), although reported rates of sexual recidivism typically refer just to sexual conduct problems severe enough to qualify as new sexual offenses. Even including sexual conduct problems less severe than sexual reoffending, however, the percentage of posttreatment sexual conduct problems in our sample may still have been low enough in the absolute sense to make it difficult to ascertain what factors potentially influence this outcome. Third, we identified pretreatment psychopathology in our sample on the basis of clinician judgment. Objective measures of psychological functioning might certainly prove to be more closely related to treatment outcome. Fourth, the treatment program in which these adolescents were enrolled is relatively intensive, long-term, and expensive. These qualities may not generalize to the average adolescent treatment program, and our sample may also have had related characteristics (e.g., relatively high socioeconomic status of their family-of-origin) that may negatively impact the generalizability of our results. Finally, and perhaps most importantly, the OQ-45 can be used to measure the trajectory of subjective distress during treatment, but perhaps reductions in such distress is not the appropriate measure of psychotherapeutic success in this population. It is here that the ASO treatment literature as a whole is still in its infancy. When ASO treatment is successful, what are the mechanisms of change, and how might we best go about measuring them?

Adolescents who have offended sexually require treatment, for reasons related both to public safety and the fact that these adolescents deserve treatment. Treatment outcome research with this population, however, is still in its infancy. The present study examined the relationship among a variety of individual differences, posttreatment outcomes, and subjective distress over the course of long-term residential treatment.

These results did not provide support for using subjective distress to anticipate negative posttreatment outcomes, but we did uncover several pretreatment characteristics that impact how adolescents fare over the course of treatment. These findings deserve further consideration; working to understand how to help all adolescents benefit fully from treatment is as worthy a goal as the prevention of new sexual offenses.

REFERENCES

Abel, G. G. (1995). *Abel Screening System Manual*. Atlanta, GA: Abel Screening.

Abel, G. G., Jordan, A., Rouleau, J. L., Emerick, R., Barboza-Whitehead, S., & Osborn, C. (2004). Use of visual reaction time to assess male adolescents who molest children. *Sexual Abuse: Journal of Research and Treatment, 16*(3), 255-265. doi: 10.1177/107906320401600306

Abel, G. G., Osborn, C. A., & Twigg, D. A. (1993). Sexual assault through the life span: Adult offenders with juvenile histories. In H. E. Barbaree, W. L. Marshall & S. M. Hudson (Eds.), *The juvenile sex offender* (pp. 104-117). New York, NY: Guilford Press.

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

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39-55. doi: 10.1037/a0018362

Atkins, D. C., Berns, S. B., George, W. H., Doss, B. D., Gattis, K., & Christensen, A. (2005). Prediction of response to treatment in a randomized clinical trial of marital therapy. *Journal of Consulting and Clinical Psychology, 73*(5), 893-903. doi: 10.1037/0022-006x.73.5.893

Becker, J. V., Hunter, J. A., Goodwin, D., Kaplan, M. S., & Martinez, D. (1992). Test-retest reliability of audio-taped phallometric stimuli with adolescent sexual offenders. *Sex Abuse, 5*(45), 45-51. doi: 10.1007/bf00849876

Becker, J. V., Kaplan, M. S., & Tenke, C. E. (1992). The relationship of abuse history, denial and erectile response profiles of adolescent sexual perpetrators. *Behavior Therapy, 23*(1), 87-97. doi: 10.1016/s0005-7894(05)80310-7

Blanchard, R., & Barbaree, H. E. (2005). The strength of sexual arousal as a function of the age of the sex offender: Comparisons among pedophiles, hebephiles, and teleiophiles. *Sexual Abuse: Journal of Research and Treatment, 17*(4), 441-456. doi: 10.1177/107906320501700407

Blanchard, R., Klassen, P., Dickey, R., Kuban, M. E., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13*(1), 118-126.

Borduin, C. M., Henggeler, S. W., Blaske, D. M., & Stein, R. J. (1990). Multisystemic treatment of adolescent sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 34*(2), 105-113. doi: 10.1177/0306624x9003400204

Borduin, C. M., Schaeffer, C. M., & Heiblum, N. (2009). A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: Effects on youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology, 77*(1), 26-37. doi: 10.1037/a0013035

Boutwell, B. B., Barnes, J. C., & Beaver, K. M. (2013). Life-course persistent offenders and the propensity to commit sexual assault. *Sexual Abuse: A Journal of Research and Treatment, 25*(1), 69-81. doi: 10.1177/1079063212452616

Bremer, J. F. (1998). Challenges in the assessment and treatment of sexually abusive adolescents. *The Irish Journal of Psychology, 19*(1), 82-92.

Burton, D. L. (2000). Were adolescent sexual offenders children with sexual behavior problems? *Sexual Abuse: Journal of Research and Treatment, 12*(1), 37-48. doi: 10.1177/107906320001200105

Caldwell, M. F. (2010). Study characteristics and recidivism base rates in juvenile sex offender recidivism. *International Journal of Offender Therapy and Comparative Criminology, 54*(2), 197-212. doi: 10.1177/0306624x08330016

Caldwell, M. F. (2013). Accuracy of sexually violent person assessments of juveniles adjudicated for sexual offenses. *Sexual Abuse: A Journal of Research and Treatment*. doi: 10.1177/1079063213480818

Carpentier, J., & Proulx, J. (2011). Correlates of recidivism among adolescents who have sexually offended. *Sexual Abuse: Journal of Research and Treatment, 23*(4), 434-455. doi: 10.1177/1079063211409950

Clift, R. J. W., Rajlic, G., & Gretton, H. M. (2009). Discriminative and predictive validity of the penile plethysmograph in adolescent sex offenders. *Sexual Abuse: Journal of Research and Treatment, 21*(3), 335-362. doi: 10.1177/1079063209338491

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Categorical of nominal independent variables. In J. Cohen, P. Cohen, S. G. West & L. S. Aiken (Eds.), *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed., pp. 302-353). New York, NY: Routledge.

Cooper, H. M. (2000). Long-term follow-up of community-based treatment program for adolescent sex offenders. *Masters Abstracts International, 45*(3).

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*(1), 583-619. doi: doi:10.1146/annurev.psych.093008.100356

Dennison, S., & Leclerc, B. (2011). Developmental factors in adolescent child sexual offenders: A comparison of nonrepeat and repeat sexual offenders. *Criminal Justice and Behavior, 38*(11), 1089-1102. doi: 10.1177/0093854811417076

DiCataldo, F. C. (2009). *The perversion of youth: Controversies in the assessment and treatment of juvenile sex offenders*. New York, NY: New York University Press.

Eastman, B. J. (2005). Variables associated with treatment failure among adolescent sex offenders. *Journal of Offender Rehabilitation, 42*(3), 23-40. doi: 10.1300/J076v42n03_02

Edwards, R., Beech, A., Bishopp, D., Erikson, M., Friendship, C., & Charlesworth, L. (2005). Predicting dropout from a residential programme for adolescent sexual abusers using pre-treatment variables and implications for recidivism. *Journal of Sexual Aggression, 11*(2), 139-155. doi: 10.1080/13552600500063641

Ellsworth, J. R., Lambert, M. J., & Johnson, J. (2006). A comparison of the Outcome Questionnaire-45 and Outcome Questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology & Psychotherapy, 13*(6), 380-391. doi: 10.1002/cpp.503

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138. doi: 10.1037/1082-989X.12.2.121.supp (Supplemental)

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: The statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology & Psychotherapy, 8*(4), 231-242. doi: 10.1002/cpp.286

Finkelhor, D., Ormrod, R., & Chaffin, M. (2009). *Juveniles who commit sex offenses against minors*. Washington, DC: Office of Justice Programs.

Fischer, L., & Smith, G. (1999). Statistical adequacy of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: Journal of Research and Treatment, 11*(3), 195-205. doi: 10.1177/107906329901100303

Gerhold, C. K., Browne, K. D., & Beckett, R. (2007). Predicting recidivism in adolescent sexual offenders. *Aggression and Violent Behavior, 12*(4), 427-438. doi: 10.1016/j.avb.2006.10.004

Glasgow, D. V. (2009). Affinity: The development of a self-report assessment of paedophile sexual interest incorporating a viewing time validity measure. In D. Thornton & D. R. Laws (Eds.), *Cognitive approaches to the assessment of sexual interest in sexual offender* (pp. 59-84). Hoboken, NJ: Wiley-Blackwell.

Gretton, H. M., McBride, M., Hare, R. D., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior, 28*(4), 427-449. doi: 10.1177/009385480102800403

Guerra, N. G., Williams, K. R., Tolan, P. H., & Modecki, K. L. (2008). Theoretical and research advances in understanding the causes of juvenile offending. In R. D. Hoge, N. G. Guerra & P. Boxer (Eds.), *Treating the juvenile offender* (pp. 33-53). New York, NY: The Guilford Press.

Hagan, M. P., Anderson, D. L., Caldwell, M. S., & Kemper, T. S. (2010). Five-year accuracy of assessments of high risk for sexual recidivism of adolescents. *International Journal of Offender Therapy and Comparative Criminology, 54*(1), 61-70. doi: 10.1177/0306624x08325348

Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior, 36*(9), 865-891. doi: 10.1177/0093854809338545

Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the Collaborative Outcome Data Project on the Effectiveness of Psychological Treatment for Sex Offenders. *Sexual Abuse: A Journal of Research and Treatment, 14*(2), 169-194. doi: 10.1177/107906320201400207

Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*(6), 1154-1163.

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

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

Hendriks, J., & Bigleveld, C. (2008). Recidivism among juvenile sex offenders after residential treatment. *Journal of Sexual Aggression, 14*(1), 19-32.

Hunter, J. A., & Becker, J. V. (1994). The role of deviant sexual arousal in juvenile sexual offending: Etiology, evaluation, and treatment. *Criminal Justice and Behavior, 21*(1), 132-149. doi: 10.1177/0093854894021001009

Hunter, J. A., Goodwin, D. W., & Becker, J. V. (1994). The relationship between phallometrically measured deviant sexual arousal and clinical characteristics in juvenile sexual offenders. *Behaviour Research and Therapy, 32*(5), 533-538. doi: 10.1016/0005-7967(94)90142-2

Imhoff, R., Schmidt, A. F., Nordsiek, U., Luzar, C., Young, A. W., & Banse, R. (2010). Viewing time effects revisited: Prolonged response latencies for sexually attractive targets under restricted task conditions. *Archives of Sexual Behavior, 39*(6), 1275-1288. doi: 10.1007/s10508-009-9595-2

Israel, E., & Strassberg, D. S. (2009). Viewing time as an objective measure of sexual interest in heterosexual men and women. *Archives of Sexual Behavior, 38*(4), 551-558. doi: 10.1007/s10508-007-9246-4

Kaemingk, K. L., Koselka, M., Becker, J. V., & Kaplan, M. S. (1995). Age and adolescent sexual offender arousal. *Sexual Abuse: Journal of Research and Treatment, 7*(4), 249-257. doi: 10.1007/bf02256829

Kalmus, E., & Beech, A. R. (2005). Forensic assessment of sexual interest: A review. *Aggression and Violent Behavior, 10*(2), 193-217.

Kingston, D. A., Yates, P. M., & Olver, M. E. (2013). The self-regulation model of sexual offending: Intermediate outcomes and posttreatment recidivism. *Sexual Abuse: A Journal of Research and Treatment*. doi: 10.1177/1079063213495896

Kraemer, B. D., Salisbury, S. B., & Spielman, C. R. (1998). Pretreatment variables associated with treatment failure in a residential juvenile sex-offender program. *Criminal Justice and Behavior, 25*(2), 190-202. doi: 10.1177/0093854898025002003

Kuban, M., Barbaree, H. E., & Blanchard, R. (1999). A comparison of volume and circumference phallometry: Response magnitude and method agreement. *Archives of Sexual Behavior, 28*(4), 345-359. doi: 10.1023/A:1018700813140

Lambert, M. J., Hansen, N. B., & Harmon, S. C. (2010). Outcome Questionnaire System (The OQ System): Development and practical applications in healthcare settings. In M. Barkham, G. E. Hardy & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 141-154). Hoboken, NJ: Wiley-Blackwell.

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology & Psychotherapy, 9*(3), 149-164. doi: 10.1002/cpp.333

Laws, D. R., & Gress, C. L. Z. (2004). Seeing things differently: The viewing time alternative to penile plethysmography. *Legal and Criminological Psychology, 9*(2), 183-196. doi: 10.1348/1355325041719338

Letourneau, E. J., & Miner, M. H. (2005). Juvenile sex offenders: A case against the legal and clinical status quo. *Sexual Abuse: Journal of Research and Treatment, 17*(3), 293-312. doi: 10.1177/107906320501700304

Lieb, R. (1996). Community notification laws: "A step toward more effective solutions." *Journal of Interpersonal Violence, 11*(2), 298-300. doi: 10.1177/088626096011002013

Longo, R. E., & Groth, A. N. (1983). Juvenile sexual offenses in the histories of adult rapists and child molesters. *International Journal of Offender Therapy and Comparative Criminology, 27*(2), 150-155. doi: 10.1177/0306624x8302700207

Lykins, A. D., Cantor, J. M., Kuban, M. E., Blak, T., Dickey, R., Klassen, P. E., & Blanchard, R. (2010). The relation between peak response magnitudes and agreement in diagnoses obtained from two different phallometric tests for pedophilia. *Sexual Abuse: Journal of Research and Treatment, 22*(1), 42-57. doi: 10.1177/1079063209352094

Mahoney, J. M., & Strassberg, D. S. (1991). Voluntary control of male sexual arousal. *Archives of Sexual Behavior, 20*(1), 1-16. doi: 10.1007/bf01543003

McCann, K., & Lussier, P. (2008). Antisociality, sexual deviance, and sexual reoffending in juvenile sex offenders: A meta-analytical investigation. *Youth Violence and Juvenile Justice, 6*(4), 363-385. doi: 10.1177/1541204008320260

McCoach, D. B., & Black, A. C. (2008). Evaluation of model fit and adequacy. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 245-272). Charlotte, NC: Information Age Publishing.

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

Miner, M. H. (2002). Factors associated with recidivism in juveniles: An analysis of serious juvenile sex offenders. *Journal of Research in Crime and Delinquency, 39*(4), 421-436. doi: 10.1177/002242702237287

Mokros, A., Gebhard, M., Heinz, V., Marschall, R. W., Nitschke, J., Glasgow, D. V., . . . Laws, D. R. (2013). Computerized assessment of pedophilic sexual interest through self-report and viewing time: Reliability, validity, and classification accuracy of the Affinity Program. *Sexual Abuse: A Journal of Research and Treatment, 25*(3), 230-258. doi: 10.1177/1079063212454550

Mueller, R. M., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70*(2), 248-262. doi: 10.1207/s15327752jpa7002_5

Murphy, W. D., DiLillo, D., Haynes, M. R., & Steere, E. (2001). An exploration of factors related to deviant sexual arousal among juvenile sex offenders. *Sexual Abuse: Journal of Research and Treatment, 13*(2), 91-103. doi: 10.1177/107906320101300203

Parks, G. A., & Bard, D. E. (2006). Risk factors for adolescent sex offender recidivism: Evaluation of predictive factors and comparison of three groups based upon victim type. *Sexual Abuse: Journal of Research and Treatment, 18*(4), 319-342. doi: 10.1177/107906320601800402

Quinsey, V. L., Ketsetzis, M., Earls, C., & Karamanoukian, A. (1996). Viewing time as a measure of sexual interest. *Ethology & Sociobiology, 17*(5), 341-354. doi: 10.1016/s0162-3095(96)00060-x

Raftery, A. E. (1995). Bayesian Model selection in social research. *Sociological Methodology, 25*, 111-163.

Rasmussen, L. A. (1999). Factors related to recidivism among juvenile sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 11*(1), 69-85. doi: 10.1177/107906329901100106

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2011). HML 7 for Windows. Skokie, IL: Scientific Software International, Inc.

Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. high-school data. *Journal of Educational Statistics, 16*(4), 295-330. doi: 10.2307/1165105

Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin, 128*(2), 330-366. doi: 10.1037/0033-2909.128.2.330

Rice, M. E., Harris, G. T., Lang, C., & Chaplin, T. C. (2012). Adolescents who have sexually offended: Is phallometry valid? *Sexual Abuse: Journal of Research and Treatment, 24*(2), 133-152.

Rullo, J. E., Strassberg, D. S., & Israel, E. (2010). Category-specificity in sexual interest in gay men and lesbians. *Archives of Sexual Behavior, 39*(4), 874-879. doi: 10.1007/s10508-009-9497-3

Schram, D. D., & Milloy, C. D. (1995). Community notification: A study of offender characteristics and recividism. Olympia, WA: Washington State Institute for Public Policy.

Seabloom, W., Seabloom, M. E., Seabloom, E., Barron, R., & Hendrickson, S. (2003). A 14- to 24-year longitudinal study of a comprehensive sexual health model treatment program for adolescent sex offenders: Predictors of successful completion and subsequent criminal recidivism. *International Journal of Offender Therapy and Comparative Criminology, 47*(4), 468-481. doi: 10.1177/0306624x03253849

Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: Journal of Research and Treatment, 13*(1), 15-25.

Seto, M. C., & Lalumière, M. L. (2010). What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychological Bulletin, 136*(4), 526-575. doi: 10.1037/a0019700

Seto, M. C., Lalumière, M. L., & Blanchard, R. (2000). The discriminative validity of a phallometric test for pedophilic interests among adolescent sex offenders against children. *Psychological Assessment, 12*(3), 319-327.

Seto, M. C., Murphy, W. D., Page, J., & Ennis, L. (2003). Detecting anomalous sexual interests in juvenile sex offenders. *Annals of the New York Academy of Sciences, 989*, 118-130.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Smith, G., & Fischer, L. (1999). Assessment of juvenile sexual offenders: Reliability and validity of the Abel Assessment for Interest in Paraphilias. *Sexual Abuse: Journal of Research and Treatment, 11*(3), 207-216. doi: 10.1177/107906329901100304

Spice, A., Viljoen, J. L., Latzman, N. E., Scalora, M. J., & Ullman, D. (2012). Risk and protective factors for recidivism among juveniles who have offended sexually. *Sexual Abuse: A Journal of Research and Treatment*. doi: 10.1177/1079063212459086

Tong, D. (2007). The penile plethysmograph, Abel Assessment for Sexual Interest, and MSI-II: Are they speaking the same language? *American Journal of Family Therapy, 35*(3), 187-202.

U.S. Department of Health and Human Services, A. f. C. a. F., Administration on Children, Youth and Families, Children's Bureau. (2010). *Child Maltreatment 2009*. Retrieved from http://www.acf.hhs.gov/programs/cb/stats_research/index.htm#can.

van der Put, C. E., van Vugt, E. S., Stams, G. J. J. M., Deković, M., & van der Laan, P. H. (2013). Differences in the prevalence and impact of risk factors for general recidivism between different types of juveniles who have committed sexual offenses (JSOs) and juveniles who have committed nonsexual offenses (NSOs). *Sexual Abuse: A Journal of Research and Treatment, 25*(1), 41-68. doi: 10.1177/1079063212452615

Vermeersch, D. A., Whipple, J. L., Lambert, M. J., Hawkins, E. J., Burchfield, C. M., & Okiishi, J. C. (2004). Outcome questionnaire: Is it sensitive to changes in counseling center clients? *Journal of Counseling Psychology, 51*(1), 38-49. doi: 10.1037/0022-0167.51.1.38

Worling, J. R. (2006). Assessing sexual arousal with adolescent males who have offended sexually: Self-report and unobtrusively measured viewing time. *Sexual Abuse: Journal of Research and Treatment, 18*(4), 383-400. doi: 10.1177/107906320601800406

Worling, J. R., & Curwen, T. (2000). Adolescent sexual offender recidivism: Success of specialized treatment and implications for risk prediction. *Child Abuse & Neglect, 24*(7), 965-982.

Worling, J. R., & Långström, N. (2006). Risk of sexual recidivism in adolescents who offend sexually. In H. E. Barbaree & W. L. Marshall (Eds.), *The juvenile sex offender* (2nd ed., pp. 219-247). New York, NY: The Guilford Press.