

Treatment Outcome, Risk Assessment, and Recidivism among Sexual Offenders against Children

A thesis submitted in fulfillment
of the requirements for a
Doctorate of Philosophy in Psychology
at the University of Canterbury

by
Sarah M. Beggs

February 2008

Acknowledgements

There are many people who have helped make this thesis possible, and I would like to sincerely thank all of them. Firstly, my supervisor, Randy Grace, for his openness to discussing ideas, his time and effort in assisting and editing, and most of all his unwavering confidence in me and encouragement throughout this process. Secondly, my family and friends who have provided so much support along the way. In particular, my parents, Mac and Sue, who have always been amazing supporters of me, and the friends and classmates who showed me it could be done, have been there for me at various times offering help and support, and shared good times when I needed a break – Andy, Dione, Gwen, Alice, Wendy, Zoe, Aimée, Tobias, Paula, and Annmaree, to name a few.

I am extremely grateful for the financial support from the Ryoichi Sasakawa Young Leaders' Fellowship, without which this research may not have been possible. Also the University of Canterbury Research Award and the Psychology Department for additional funding for these projects; and the Department of Corrections Psychological Service, particularly the Kia Marama Special Treatment Unit, for providing access to the data, transferring files, and equipping me with space to carry out this research. Thank you to Todd Hogue, for provision of his SGAS materials, and Mark Olver for the provision of the VRS:SO manual and training on its use. I would finally like to acknowledge the hard work and commitment of the research assistants who assisted me with the data coding – Annmaree Kingi and Jo Vallance, and also sincerely thank Peter Kelly and Sandra Hicks for their time and valuable comments on portions of an earlier draft of this manuscript.

Author Note

Portions of the data reported here have previously been presented at the annual meetings of the Australian and New Zealand Association for the Treatment of Sexual Abusers, the New Zealand Psychological Society, the New Zealand College of Clinical Psychologists, the Association for the Treatment of Sexual Abusers, and the International Association for the Treatment of Sexual Offenders.

Study 1 of this dissertation has been accepted pending minor revisions for publication in *Sexual Abuse: A Journal of Research and Treatment*, and Study 4 is currently in press in *Criminal Justice and Behavior*.

Note that U.K. spelling is used in this manuscript except for the empirical section, in which U.S. spelling has been used in preparation for submission to international journals.

Table of Contents

Abstract.....	1
Literature Review.....	4
<i>Overall Introduction and Outline of Literature Review.....</i>	4
<i>Treatment of Child Sex Offenders – Overview.....</i>	8
<i>Brief Historical Development of Sex Offender Treatment Programmes.....</i>	8
<i>Current Practice of Sex Offender Treatment.....</i>	9
<i>Theoretical frameworks for treatment.</i>	10
<i>Format and processes.</i>	14
<i>Components and content.</i>	15
<i>Kia Marama – An example of a treatment programme for child sexual offenders. ..</i>	19
<i>Does Treatment for Sex Offenders Work?.....</i>	22
<i>Risk Assessment Among Sex Offenders.....</i>	30
<i>First Generation of Offender Risk Assessment: Unstructured Clinical Judgement.....</i>	30
<i>Risk Factors – Static and Dynamic.....</i>	31
<i>Second Generation of Offender Risk Assessments: Actuarial Risk Instruments.....</i>	32
<i>Third Generation Risk-Needs Assessments: Incorporation of Dynamic Factors.....</i>	36
<i>Fourth Generation? New Directions in Risk Assessment.....</i>	42
<i>Treatment Outcome – Conceptualisations and Relationship with Recidivism.....</i>	47
<i>Why Measure Treatment Outcome?.....</i>	47
<i>Conceptualisations of Treatment Outcome – Literature Search.....</i>	49
<i>Change on dynamic risk.</i>	59
<i>Risk assessment tools incorporating treatment outcome.</i>	64
<i>Systematic clinical ratings of overall treatment outcome.</i>	68
<i>Summary.....</i>	74
<i>Outline of the Empirical Section.....</i>	77
Study 1. Assessment of Dynamic Risk Factors: An Independent Validation Study of the Violence Risk Scale: Sexual Offender Version.....	82
1.1 Abstract.....	82
1.2 Introduction.....	83
1.3 Method.....	87
1.3.1 Participants.....	87
1.3.2 Measures and Data Collection Procedures.....	87
1.3.2.1 VRS:SO.	87
1.3.2.2 Static-99.	89
1.3.2.3 Recidivism.	89
1.3.2.4 Psychometric measures of dynamic risk factors.	90
1.3.3 Data Analyses.....	91
1.4 Results.....	92
1.4.1 Inter-Rater Reliability.....	92
1.4.2 Factor Analyses.....	92
1.4.3 Descriptive Statistics and Offender Group Comparisons.....	95

1.4.4	<i>Concurrent Validity</i>	96
1.4.5	<i>Predictive Validity</i>	97
1.4.6	<i>Risk Categories</i>	99
1.4.7	<i>Contributions of Dynamic Risk Controlling for Static Risk</i>	101
1.4.8	<i>VRS:SO and Psychometric Measures of Dynamic Risk</i>	102
1.5	<i>Discussion</i>	103
Study 2. Measuring Treatment Outcome for Child Molesters – A Comparative Validity		
	Study.....	108
2.1	<i>Abstract</i>	108
2.2	<i>Introduction</i>	109
2.3	<i>Method</i>	114
2.3.1	<i>Participants</i>	114
2.3.2	<i>Measures of Treatment Outcome</i>	115
2.3.2.1	<i>Psychometric change</i>	116
2.3.2.2	<i>Violence Risk Scale: Sexual Offender Version change scores</i>	117
2.3.2.3	<i>Standard Goal Attainment Scaling</i>	118
2.3.3	<i>Static Risk Measure</i>	119
2.3.4	<i>Recidivism</i>	119
2.4	<i>Results</i>	119
2.4.1	<i>Psychometric Change</i>	119
2.4.1.1	<i>Standardized residual change scores</i>	124
2.4.1.2	<i>Clinically significant and reliable change</i>	124
2.4.2	<i>Measures of Treatment Outcome Based on File Review – VRS:SO and SGAS</i> ...	128
2.4.3	<i>Relationship Between Treatment Outcome and Recidivism</i>	130
2.4.3.1	<i>Psychometric change and recidivism</i>	131
2.4.3.2	<i>Violence Risk Scale: Sexual Offender Version and recidivism</i>	135
2.4.3.3	<i>Standard Goal Attainment Scaling and recidivism</i>	136
2.4.4	<i>Can Measures of Treatment Outcome Contribute to Risk Prediction Above and Beyond Static Factors?</i>	137
2.5	<i>Discussion</i>	139
Study 3. Psychopathy, Treatment Goal Attainment, and Sex Offender Recidivism.....		
3.1	<i>Abstract</i>	145
3.2	<i>Introduction</i>	146
3.3	<i>Method</i>	151
3.3.1	<i>Participants</i>	151
3.3.2	<i>Measures</i>	151
3.3.2.1	<i>Psychopathy</i>	151
3.3.2.2	<i>Treatment outcome</i>	152
3.3.2.3	<i>Recidivism</i>	152
3.4	<i>Results</i>	153
3.4.1	<i>Demographic and Offense Information</i>	153
3.4.2	<i>Psychopathy and Treatment Outcome</i>	153
3.4.3	<i>Recidivism</i>	155
3.4.3.1	<i>Survival analyses</i>	158
3.4.3.2	<i>Median-split analyses</i>	160
3.4.4	<i>Incest and Extrafamilial Offenders Compared</i>	163
3.4.4.1	<i>Survival analyses</i>	164
3.4.4.2	<i>Median-split analyses</i>	165

3.5 Discussion.....	166
Study 4. Intelligence and Psychopathy – Evidence for an Interaction Effect on Recidivism of Child Sexual Offenders.....	172
4.1 Abstract.....	172
4.2 Introduction.....	173
4.3 Method.....	177
4.3.1 Participants.....	177
4.3.2 Measures.....	177
4.3.2.1 Psychopathy.....	177
4.3.2.2 Intelligence.....	178
4.3.2.3 Actuarial risk.....	178
4.3.2.4 Offending.....	179
4.4 Results.....	179
4.4.1 Demographic and Offense Information.....	179
4.4.2 Psychopathy, Intelligence, and Actuarial Risk.....	181
4.4.3 Offending.....	182
4.4.3.1 Is there an interaction effect between psychopathy and intelligence on offending?.....	183
4.4.3.2 Is the interaction specific to psychopathy?.....	185
4.4.3.3 Predictive validity.....	186
4.4.3.4 Incest and extrafamilial comparisons.....	187
4.5 Discussion.....	189
General Discussion.....	193
Overview of the Empirical Results of this Dissertation.....	193
Concluding Discussion – Key Themes.....	199
References.....	203

List of Tables

Table 1	<i>Review of Studies Incorporating Assessment of Treatment Outcome.....</i>	51
Table 2	<i>Four-Factor Loading Matrix for VRS:SO Dynamic Items.....</i>	94
Table 3	<i>Means and Standard Deviations of Static-99 and VRS:SO Scale Scores for Incest Offenders, Extrafamilial Offenders, and the Total Sample.....</i>	95
Table 4	<i>Predictive Accuracy of the Static-99, and VRS:SO Scale Scores (Pre- and Post-Treatment) for Sexual, Violent, and General Recidivism.....</i>	98
Table 5	<i>Correlations Between Factor Scores Derived From Psychometric Self-Reports by Allan et al. (2007) and VRS-SO Dynamic Risk Scores and Sexual Recidivism.....</i>	103
Table 6	<i>Means, Standard Deviations, and N for Pre-Treatment and Post-Treatment Psychometric Variables, and Treatment Change Effect Sizes (Cohen's <i>d</i>).....</i>	121
Table 7	<i>Percentage that Achieved Clinically Significant Change (CSC), Percentage Achieving CSC who were Deviant Pre-Treatment, and Percentage that Achieved Both Reliable and Clinically Significant Change (RCSC) on Each Test.....</i>	127
Table 8	<i>Means and Standard Deviations of VRS:SO Change Scores (N = 218), and SGAS Scores (N = 223).....</i>	128
Table 9	<i>Correlations Between Measures of Treatment Outcome: Standardized Residual Change (RCZ), Clinically Significant Change (CSC) and Reliable and Clinically Significant Change (RCSC) on the Four-Factor Psychometric Battery; Standard Goal Attainment Scaling (SGAS) Total Scores; and Change Scores on the Violence Risk Scale: Sexual Offender Version (VRS:SO) Total Dynamic scale and Three Factors.....</i>	129
Table 10	<i>Correlations Between Raw Psychometric Change Scores and Sexual Recidivism, and Partial Correlations (Controlling for Pre-Treatment Scores).....</i>	132
Table 11	<i>Predictive Validity of Measures of Treatment Outcome: Correlations with Sexual Recidivism and AUCs for Change on a Four-Factor Psychometric Battery (RCZ, CSC and RCSC); SGAS Scores; and Change on VRS:SO Dynamic Scale and Factor Scores.....</i>	134
Table 12	<i>Logistic Regression Analyses of the Predictive Validity of Measures of Treatment Outcome (RCZ, CSC, SGAS, VRS:SO change) Controlling for Static-99.....</i>	138
Table 13	<i>Means and Standard Deviations of Age, PCL-R Scores and SGAS Scores.....</i>	154
Table 14	<i>Correlations – Demographic Variables, PCL-R, and SGAS.....</i>	156
Table 15	<i>Correlations – Demographic Variables, PCL-R, SGAS, and Recidivism.....</i>	157
Table 16	<i>Cox Regression Results for Tests of PCL-R, SGAS, and Interaction on Recidivism.....</i>	159
Table 17	<i>Group Sizes and Recidivism Rates Following Median-Splits of PCL-R (High/Low) and Total SGAS (Good/Poor) Scores.....</i>	160
Table 18	<i>Means and Standard Deviations for Test Scores, Prior Convictions, and Recidivism.....</i>	179
Table 19	<i>Correlations Between PCL-R, IQ, and Static-99 Scores and Offense Variables.....</i>	182
Table 20	<i>Group Sizes and Recidivism Rates of the Total Sample and Four Groups Following Median-Splits of IQ and PCL-R Scores.....</i>	184
Table 21	<i>Group Sizes and Recidivism Rates of the Total Sample and Four Groups Following Median-Splits of IQ and Static-99 Scores.....</i>	186
Table 22	<i>Areas Under the Receiver Operating Characteristic Curve (AUC) for the PCL-R and Static-99 Predicting Sexual, Violent, and General Recidivism.....</i>	187
Table 23	<i>Group Sizes and Recidivism Rates of the Total Sample and Four Groups Following Median-Splits of IQ and PCL-R Scores.....</i>	188

List of Figures

<i>Figure 1</i>	Recidivism base rates for risk categories derived from Total VRS:SO scores.....	100
<i>Figure 2</i>	Kaplan-Meier survival plot showing cumulative sexual recidivism failure rates as a function of VRS:SO risk level.....	101
<i>Figure 3</i>	Change scores for the ABCS as a function of pre-treatment score.....	123
<i>Figure 4</i>	Kaplan-Meier survival plot for serious recidivism for PCL-R and Total SGAS median-split groups.....	161
<i>Figure 5</i>	Kaplan-Meier survival plot for sexual recidivism for PCL-R and Total SGAS median-split groups.....	162

Abstract

The sexual abuse of children is an issue that society must address with urgency and commitment, given the profoundly damaging effects and widespread occurrence of this kind of crime. Providing psychological treatment to identified offenders is an important endeavour of the criminal justice system, with the aim of reducing recidivism and thereby preventing future victims. This dissertation explores a number of areas relevant to the treatment of sexual offenders on a sample of 223 adult males who completed a prison-based programme for child sexual offenders in New Zealand. Specifically, the assessment of treatment outcome and its relationship with recidivism, risk assessment, and the influence of specific offender factors on estimates of treatment outcome and risk were investigated. Study 1 ($N = 218$) is an independent validation of the validity of the Violence Risk Scale: Sexual Offender Version (VRS:SO; Olver, Wong, Nicholaichuk, & Gordon, 2007), a recently developed risk assessment instrument for sexual offenders that incorporates both static and dynamic risk factors and contains protocols for the assessment of change as a result of treatment. Results indicate support for the inter-rater reliability, concurrent validity, and predictive validity of the VRS:SO with regard to sexual recidivism, with pre-treatment and post-treatment scores showing superior predictive validity relative to a widely used measure of static risk (Static-99; Hanson & Thornton, 1999) and a measure of “Deviance” based on a 4-factor battery of relevant psychometric tests (Allan, Grace, Rutherford, & Hudson, 2007).

In Study 2 ($N = 218$), three separate methods of assessing proximal treatment outcome (representative of three categories of treatment outcome measures that have

previously been applied in the literature) are applied and compared in terms of their predictive validity with regard to sexual recidivism, and the relative advantages and disadvantages of their use. These measures are: change on a battery of relevant psychometric tests administered prior to and following treatment; change across treatment on the VRS:SO; and post-treatment ratings of the attainment of treatment goals as measured by a modified version of Hogue's (1994) Standard Goal Attainment Scaling for Sexual Offenders (SGAS). Results indicate that positive treatment outcomes as measured by all of these methods are associated with reduced sexual recidivism. SGAS scores are identified as being relatively simple and efficient to obtain, however the VRS:SO and the psychometric battery are both able to provide useful pre-treatment clinical information regarding potential treatment targets for a particular offender.

Study 3 ($N = 223$) and Study 4 ($N = 216$) are explorations of the influence of particular offender characteristics on response to treatment and risk of recidivism. Of particular interest was the personality construct of psychopathy (measured using the Psychopathy Checklist-Revised, PCL-R; Hare, 1991), and both studies are attempted replications and extensions of previously reported interaction effects involving this construct (Heilbrun, 1979; Seto & Barbaree, 1999). The results of Study 3 indicate that there is no interaction effect between PCL-R scores and treatment outcome (as measured by the SGAS) on sexual recidivism, in contrast to an influential study by Seto and Barbaree (1999). Study 4 reports an interaction effect between PCL-R scores and intelligence on recidivism, such that higher than average IQ scores appear to moderate the well-known association between psychopathy and risk. Overall, the findings reported in this dissertation suggest the importance of considering dynamic factors as well as static factors in sex offender risk assessments, and support the

premise that dynamic factors are changeable, with change being associated with changes in recidivism. The potential for certain offender characteristics to influence treatment response and risk of recidivism is highlighted, and several areas for further exploration are identified.

Literature Review

Overall Introduction and Outline of Literature Review

Sexual offending is a major concern for societies, due to the prevalence of this kind of crime and the serious negative consequences for victims. Although legal definitions vary by jurisdiction, Shaw (2004) offered the following general description of sexually abusive behaviour: “any sexual behaviour which occurs: 1) without consent, 2) without equality, or 3) as a result of coercion” (p. 217). With regard to sexual abuse perpetrated against children, Shaw contended that children lack the ability to consent in any circumstances, both legally and by definition of their limited cognitive development.

Sexual crimes are known to be widely under-reported, particularly those against child victims (Marshall, Anderson, & Fernandez, 1999). Official crime data is therefore unable to accurately reflect the extent of the problem. General population surveys in which respondents are questioned about their experiences of being victimised may provide a better estimate of the true prevalence of sexual abuse (although this methodology is not without flaws; see Leventhal, 1998). In a New Zealand study, one in three women retrospectively reported being subjected to unwanted sexual experiences prior to the age of 16, however only 7% of this abuse was ever reported to the authorities (Anderson, Martin, Mullen, Romans, & Herbison, 1993). Finkelhor (1994) conducted an international review of similar studies from 20 countries, and reported estimates ranging from 7-36% of females and 3-29% of males, with the disparity thought to be influenced by methodological differences across studies (such as the level of detail and sensitivity in questioning subjects) rather than

true variation in prevalence. These figures endorse the view that child sexual abuse occurs much more frequently than official rates alone would indicate, and suggest that it is a serious problem for societies internationally (Finkelhor, 1994).

Sexual abuse of children is linked to a multitude of negative outcomes for victims, both in the short-term and long-term. Kendall-Tackett, Williams, and Finkelhor (1993) reviewed 45 studies on the impact of sexual abuse on child victims. Their findings showed that although there did not appear to be a syndrome pattern that was typical among child sexual abuse victims, “there is virtually no general domain of symptomatology that has not been associated with a history of sexual abuse” (p. 173). Symptoms of Posttraumatic Stress Disorder (including nightmares) and sexualised behaviours (such as sexualised play with dolls, and age-inappropriate sexual knowledge) were particularly frequent. Also common were poor self-esteem, anxiety and fear, withdrawn behaviour, regressive behaviour, running away, general behaviour problems, self-injurious behaviour, internalising behaviours such as depression and inhibition, and externalising behaviours such as aggression and antisociality (Kendall-Tackett et al., 1993). Sexual abuse is also known to have a “ripple effect” (Morrison, Quadara, & Boyd, 2007), in which the negative impact extends to others beyond the victim, such as their family, social circle, and community, as well as their future partners and future children. In some cases, victims of sexual abuse go on to sexually abuse others – approximately 50% of convicted sex offenders report being previously victimised themselves (Dhawan & Marshall, 1996).

The extensive and profoundly damaging effects of sexual abuse, coupled with the high numbers of victims, means that this is a problem that society must address with urgency and commitment. Primary prevention, increasing detection and reporting rates, apprehending and prosecuting offenders, and providing psychological

treatment to victims of sexual abuse are all important endeavours worthy of resources. Also important is the treatment of identified offenders – targeting the underlying causes of their sexually abusive behaviour in order to reduce their risk of reoffending, effectively preventing future victims.

Although the field is in a state of relative infancy, a large body of theoretical and empirical research has been conducted on the treatment of sexual offenders and related issues. In the sections of this review to follow, an overview will be provided of the development and current practice of sex offender treatment programmes internationally, including theoretical frameworks, format, processes, and content. Kia Marama, as an example of an established and currently operating treatment unit, and the site of the present research, will be described in some detail in this section. The question of treatment effectiveness – does treatment for sex offenders work in terms of reducing recidivism? – will also be addressed, using relevant data from evaluation studies and meta-analyses. Following from this, the next section will review the practice of risk assessment for treated sex offenders, including a discussion of static and dynamic risk, an overview of available risk assessment tools, and their validity in accurately predicting who will reoffend.

The concept of treatment outcome – how well an offender performs in treatment and the level of change achieved – has not been given as much attention by researchers as broader questions of overall effectiveness, and will be central to this dissertation. Past attempts to operationalise this concept among sex offenders will be reviewed, including studies using psychometric change scores, risk instruments, Goal Attainment Scaling, and various other idiosyncratic rating systems. Findings on the validity of these conceptualisations and their methodological problems will be discussed. In particular, this section will focus on the relationship between treatment

outcome and reoffending. Here, the question will be posed: Are offenders who show the greatest level of success in treatment less likely to reoffend following release compared to those who do not perform as well? This link is an important one, because of its intuitive sense. Such an assumption has the potential to influence post-treatment reports addressing risk, and so may influence release and supervision decisions. It is therefore important to ascertain whether this link (between positive treatment outcomes and reduced reoffending) is grounded in data. As will be shown in the review, the findings on this have been somewhat mixed.

At the end of the review will be an overview of the empirical section of this dissertation, including a description of the studies to follow and their relevance within the existing literature.

*Treatment of Child Sex Offenders - Overview**Brief Historical Development of Sex Offender Treatment Programmes*

Treatment programmes for sexual offenders as they exist today began to take shape from the 1960s onwards, as laboratory findings on behaviour and learning eventually started to be applied in practice to the remediation of human problems. Prior to this, psychoanalysis was the primary treatment approach for aberrant sexual behaviour (Marshall et al., 1999). Early behavioural approaches to treating sexual offenders tended to involve just one target, usually orientation of sexual interest (Fernandez, Shingler, & Marshall, 2006). Consistent with behavioural treatments in general, the problem behaviour (deviant sexual interests) was assumed to have been previously learned, and techniques such as shaping (differentially reinforcing successive approximations to a desired outcome, such as arousal to non-deviant stimuli) or aversion (repeated pairing of deviant images or thoughts with some unpleasant sensation such as a foul odour) were applied in an attempt to modify this learning.

From the 1970s, researchers and therapists in the field were becoming increasingly aware that sexual offending, like many other psychological problems, was much more multi-faceted and etiologically heterogeneous than originally thought, such that similar symptoms and behaviours did not always share the same underlying causal factors. Illustrating this, studies employing phallometric assessments of men who have committed sexual offences have indicated that only 13-48% exhibit deviant sexual interests (Marshall et al., 1999). Insight into the heterogeneity of sexual offending meant that it was no longer reasonable to assume that treatment based on a

single causal theory (behavioural modification of sexual preferences) would be equally effective for all offenders, and greater attention was turned to the elaboration of diverse theories regarding the development and maintenance of sexually abusive behaviours. This was an important endeavour, because explanatory and descriptive theories can inform as to treatment targets that may potentially be effective in reducing recidivism (Ward, Polaschek, & Beech, 2006). Treatment programmes consequently began to incorporate additional components such as social skills training, empathy enhancement, and cognitive restructuring. Typical components of current treatment programmes are described in more detail below.

Since that time, treatment approaches for sex offenders have continued to evolve into comprehensive programmes based on theoretical frameworks. Considerable theoretical and research advancements have been made, informing not only the “what” of treatment (therapeutic targets and components) but also the “how” and the “why.” Particularly important milestones occurred in the 1980s and 1990s, with the development of the risk-needs-responsivity offender classification framework (Andrews & Bonta, 2003), and the incorporation of the relapse prevention treatment model from the addictions literature (Laws, 1989); these models will be expanded on below. In the section to follow, the current state of the practice of sex offender treatment will be described in greater detail, including descriptions of relevant underlying theories, and the format, processes, and content of treatment.

Current Practice of Sex Offender Treatment

Policies, procedures, and funding for sex offender treatment in a particular jurisdiction can be affected by political factors: In particular, the value placed on

rehabilitation as opposed to other criminal justice system objectives such as retribution, deterrence, or incapacitation (see Andrews & Bonta, 2003) is variable. Programmes can also vary with the theoretical orientation of the treatment providers, type of offender and their risk level, and inclusion of culturally relevant features. Despite this variation, the last few decades of research have led to an increased consensus as to what constitutes best practice, with programmes continuing to evolve as further advancements are made. What is described below is an overview of current best practice and typical common features of modern sex offender treatment programmes.

Theoretical frameworks for treatment.

Cognitive-behavioural therapy is now generally accepted as the most effective approach currently available for treating sex offenders (Hanson et al., 2002; Ward et al., 2006). This broad label encompasses therapeutic efforts based on behavioural, social learning, and cognitive etiological perspectives of sexual offending, and treatment approaches are influenced considerably by a framework known as the “Psychology of Criminal Conduct” (PCC; Andrews & Bonta, 2003). The PCC is not specific to sexual offending, but is a general theory of the origins of criminal behaviour with an emphasis on complex individual differences viewed within social, political, economic, and historical contexts.

The PCC also provides a framework for classifying and treating offenders according to four principles of effective rehabilitation – risk, need, responsivity, and professional discretion (Andrews, Bonta, & Hoge, 1990). The risk principle states that treatment provision should be proportional to the risk level of the offender, such that higher risk offenders will derive the most benefit from intensive programmes, but

delivering intensive treatment to lower risk offenders may be ineffective or harmful. According to the need principle, treatment should target the features of offenders that are related in a causal way to their offending – these are termed criminogenic needs or dynamic risk factors. According to this model, treatment for offender needs that are not criminogenic in nature will be ineffective in reducing recidivism, traditionally the ultimate goal of treatment. The responsivity principle refers to delivering treatment in ways that maximise efficacy. General responsivity advocates structured cognitive-behavioural interventions, due to evidence suggesting better efficacy over other treatment styles (such as insight-oriented therapy, or therapeutic communities; Lösel & Schmucker, 2005). Specific responsivity refers to delivering treatment in ways consistent with the cognitive ability, learning style, personality profile, and other characteristics of the offender. External to the offender, factors such as therapist characteristics and group atmosphere are also relevant to this principle. Finally, the principle of professional discretion acknowledges that while these principles provide an empirical framework for assessment and treatment, there will be a small number of unique cases requiring special handling according to professional judgement, and so the first three principles need not be applied rigidly (Andrews et al., 1990).

Another treatment approach that has heavily influenced modern sex offender programmes is Relapse Prevention (RP), adapted from the addictions treatment literature in the 1980s (e.g., Laws, 1989). RP is a cognitive-behavioural, self-management approach that can be applied as a programme component relevant to the maintenance of treatment gains, or as an underlying framework for treatment as a whole (Dowden, Antonowicz, & Andrews, 2003). The RP model as originally applied described sexual reoffending behaviour as part of a sequential chain of events, including: decisions leading to the offender entering a high risk situation, maladaptive

or nonexistent coping responses to that situation, a behavioural lapse (i.e., offence precursor activities), and finally, a potential relapse (reoffence) depending on attributional processes following the lapse (Laws, 2003). The model has undergone some developments over time as a result of theoretical advancements (Laws, Hudson, & Ward, 2000), including a refocus on individualised offence pathways to extend the treatment model to a wider range of offenders (Hudson & Ward, 1996). However the basic premise of this approach remains that an offence chain can be broken at any point prior to the reoffence, and treatment focuses on helping the offender to understand his chain, identify and avoid his own high risk situations, and learn and apply adaptive coping strategies. Although the model is essentially based on self-management, there is also an emphasis on developing supports and encouraging openness and external monitoring.

Although these two models (risk-needs-responsivity and RP) have played major roles in shaping modern sex offender programmes, they are not without limitations, and theoretical developments are continuing to advance the practice of treatment. Specifically, the original RP model has been criticised for failing to take into account the heterogeneity of sexual offenders, its application being limited to those who are already motivated to avoid reoffending (Hanson, 2000). For example, it would seem unlikely that offenders who follow approach goal pathways as described by Ward and Hudson (1998) would adhere to treatment principles based on self-management of risk, although Laws (2003) argued that the treatment elements of RP remain useful. A further criticism is that RP treatment as it is presented to sexual offenders has an over-emphasis on negative, avoidance-based strategies that are unlikely to be motivating, and may foster resistance or even rebellion in some offenders (Mann, 2000). A “one-size fits all” approach neglects the importance of

contextual treatment factors such as the therapeutic relationship, which (as discussed in more depth below) is important to treatment efficacy (Ward & Gannon, 2006).

Similarly, Ward and Stewart (2003) have criticised the risk-needs-responsivity model for its exclusive focus on addressing criminogenic needs (i.e., dynamic risk factors) at the expense of all the other factors that are related to an individual's wellbeing and quality of life (i.e., non-criminogenic needs). They also argued that the issue of treatment responsiveness is not adequately elaborated on in the model.

An emerging model that attempts to address these limitations, while retaining a focus on reducing recidivism, is the Good Lives Model (GLM), essentially a strengths-based approach (Ward & Stewart, 2003). The basic underlying premise of the GLM is that offenders, like all humans, have a set of primary needs or "goods," such as the need for healthy functioning, autonomy, self-directedness, mastery, friendship, community, and knowledge. Offending behaviour (including sexual offending) can be viewed as inappropriate or problematic means to meet these needs. The focus of treatment based on a good lives approach therefore involves the identification of salient needs or goals of an offender (not only those that are criminogenic), the function that offending served for the individual in terms of primary goods associated with the behaviours, and the development of a "good lives plan" centred on the goals that are motivating to the offender and involving appropriate or prosocial means to meet them (e.g., skills, opportunities, values, and social supports), taking into account the context in which they will be living (Ward & Gannon, 2006). Proponents of this model have argued that maximising an offender's capacity and opportunity to achieve a good life will as a matter of course also address their criminogenic needs, and thereby reduce the risk of reoffending. The purpose of treatment therefore remains to reduce recidivism, but the therapeutic processes are

more positively focussed, encompassing approach goals which are likely to be more motivating for an offender than a focus on avoidance (Ward, Vess, Collie, & Gannon, 2006).

Format and processes.

The way that treatment is delivered is influenced by the responsivity principle. In terms of format, group therapy for sexual offenders is commonly advocated as holding several advantages over individual treatment (Spencer, 1998; Marshall et al., 1999), including increased efficiency and effectiveness. The group format allows for both direct and vicarious exposure to treatment targets, and can contribute to the process of interpersonal skill development among offenders. Group members can become involved in each others' therapy and learn from each other, offering constructive challenges that are mutually beneficial. Mutual support both within and outside of therapy hours is also available to group members.

The issue of therapeutic process was relatively neglected in the literature during the development of the cognitive-behavioural approach to treating sex offenders, in favour of more prescribed therapeutic procedures for facilitating change in thoughts, attitudes, and behaviours (Serran, Fernandez, Marshall, & Mann, 2003). Specifically, highly manualised, predominantly psychoeducational programmes were widely adopted, based on the body of empirical findings looking at "what works" with offender populations (Marshall & Serran, 2004). However, in the past 5-10 years there has been an increased focus on the contribution of therapist features and therapeutic climate to outcomes of treatment (e.g., Beech & Hamilton-Giachritsis, 2005; Marshall et al., 2002; Marshall et al., 2003). Findings that have emerged from this research indicate that therapist characteristics conducive of beneficial change on

treatment targets include empathy, warmth, rewardingness, and directiveness (Marshall et al., 2002). Conversely, a confrontational approach (defined as aggressive or derogatory challenges to denial or defensiveness) was negatively correlated with behavioural change (Marshall et al., 2003). Marshall (2005) suggested that flexibility would be another therapist characteristic important for fostering change, however this variable was not able to be reliably measured in preliminary work on this issue due to the predominance of the highly manualised approach. In this context, flexibility refers to the therapist's ability to adapt their approach to the needs of each individual client and how they are feeling in each session. Marshall, Marshall, Serran, and Fernandez (2006) reported that positive therapist features together account for between 40-60% of the variance in treatment gains, and are also conducive to a positive group climate. Features of the group climate have also been shown to predict treatment-induced benefits among sex offenders; these include group cohesiveness, and degree of emotional expression in treatment (Beech & Hamilton-Giachritsis, 2005).

Components and content.

The content of sex offender treatment is traditionally influenced by the need principle, which states that treatment targets should be based on criminogenic needs (also called dynamic risk factors) – those offender variables that are linked to sexual offending behaviours and that are theoretically associated with reductions in recidivism when changed through intervention. The specific targets addressed by programmes are therefore informed by etiological theories and empirical research on the relationship between various offender factors and recidivism.

As previously noted, early notions regarding the cause of sexual offending behaviours primarily involved deviant sexual interests (presumed to have been

learned). Although etiology is now considered heterogeneous and treatment content has expanded considerably to include additional components, the validity of deviant sexual interests as a treatment target has continued to receive support (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005). Clinically, this component usually involves behavioural interventions aimed at decreasing deviant sexual arousal and increasing appropriate or pro-social arousal. Specific techniques, described in greater detail by Marshall et al. (2006) and Hudson et al. (1998), include aversion (described above), covert sensitisation/association (repeated mental association of an offence build-up fantasy sequence with progressively earlier imagined negative consequences), masturbatory reconditioning/directed masturbation (after becoming aroused, masturbating to individually developed fantasies involving appropriate behaviours and partners), or satiation (involving repeated verbalisation of deviant fantasies in the absence of reinforcement, i.e., during the refractory period immediately following orgasm).

Cognitive distortions are another typical target in multi-component treatment programmes. Although not easily defined, Marshall et al. (2006) suggested that cognitive distortions are best understood as self-protective representations of events, actions, or people. In the case of sexual offenders these can include denial, minimisations, misperceptions of victim behaviour, rationalisations, justifications, and attitudes of entitlement or hostility towards women, and are thought to be manifestations of underlying schema (Marshall et al., 1999). Cognitive distortions (although not denial) have been broadly shown to be significantly related to sexual recidivism (Hanson & Bussière, 1998; Hanson & Harris, 2000). As a treatment target, procedures to address cognitive distortions usually involve cognitive restructuring (education regarding the role of distortions in their offending, training for how to

identify their own distortions, and challenging these distortions) and reality-testing of distorted statements with other group members, with the therapist providing corrective feedback when required (Marshall et al., 1999).

A related treatment target is empathy enhancement. This is included as a component in the majority of sex offender treatment programmes, despite both theory and research being underdeveloped on this concept (Ward et al., 2006). Indeed, empathy for victims was found to be unrelated to recidivism in a large-scale meta-analysis (Hanson & Bussière, 1998). The rationale for including an empathy component in practice therefore appears to be predominantly based on intuition; specifically, the view that the experience of empathy will impel change in distorted attitudes and inhibit potential future offending (Marshall et al., 1999). Empathy has been defined as a four-stage process involving the ability to recognise emotional states in others, take the perspective of others, and vicariously replicate these emotional states; with the fourth stage being the decision to act or not on the basis of those feelings (Marshall, Hudson, Jones, & Fernandez, 1995). Examples of treatment procedures include psychoeducation relating to victim impact in general, perspective-taking exercises such as roleplays relating to their specific victims or autobiographical accounts written as if by their victim, and the construction of mock apology letters.

There is evidence that impaired capacity for adaptive emotion regulation is associated with sexual offending as a dynamic factor, making it a potential target for treatment. Hudson, Ward, and McCormack (1999) found that 47 out of 86 sex offenders (54.7%) followed an offence pathway that was precipitated by some form of negative affect such as depression, boredom, loneliness, anger, stress or anxiety; sometimes in the context of life events such as relationship break-ups, rejection, or

financial difficulties. Additionally, Hanson and Harris (2000) found that notable increases in dysphoric moods (particularly anger) during periods of supervision were significantly predictive of recidivism in a sample of sex offenders on community supervision. In the absence of adaptive coping strategies, sex offenders are hypothesised to rely on various sexualised coping mechanisms (Marshall & Marshall, 2000). Treatment techniques are typically based on teaching offenders how to recognise their negative mood states and the link between these and their offending; and equipping them with adaptive coping strategies including relaxation, behavioural activation, and thought monitoring, as well as communication and conflict resolution skills (e.g., Hudson et al., 1998).

Intimacy deficits have been identified as a significant predictor of recidivism in large-scale meta-analysis (Hanson & Morton-Bourgon, 2005). Theoretically, it has been posited that intimacy deficits in sex offenders are a result of insecure attachment styles, developed in early childhood and resulting in lifelong inability to build secure attachment bonds with others via impaired self-esteem and social skills deficits (Burk & Burkhart, 2003). Poor understanding of intimacy and limited means to meet their needs in appropriate peer-aged relations, lead sexual offenders to turn to abusive behaviour in a distorted attempt to seek intimacy (Mulloy & Marshall, 1999). Treatment approaches to enhance offenders' capacity for intimate adult relationships may include: psychoeducation and informal discussion of research findings relating to issues in romantic relationships (such as gender differences, and the benefits of open communication, mutuality, and supportiveness); identification of unhelpful relationship styles group members may have demonstrated in the past; and group brainstorming and roleplaying of skills such as conflict resolution (Hudson, Marshall, Ward, Johnston, & Jones, 1995; Mulloy & Marshall, 1999). As the goal is to facilitate

movement towards a more secure attachment style, the enhancement of self-esteem may also be a target (e.g., Marshall et al., 2006).

Additionally, although relapse prevention concepts can form the underlying theoretical framework for treatment programmes (as discussed), they can also be included as a separate treatment component aimed at maximising the maintenance of treatment gains. Although the evidence is not supportive of the efficacy of RP treatment concepts, particularly those based on extensive external supervisory components (Marshall & Anderson, 1996), including a treatment target based on enhancing strategies for internal self-management would seem to be important for at least some sexual offenders. Self-regulation deficits were found to be a feature of the offence pathway for 49% of sexual offenders (i.e., those whose offending tended to follow avoidant-goal or approach-automatic pathways; Yates & Kingston, 2006). General self-regulation problems were also identified in Hanson and Morton-Bourgon's (2005) meta-analysis as being a potentially important dynamic risk factor, with a small to medium mean effect size with regard to recidivism of $d = .37$. Typical treatment approaches include the identification of high risk situations and the warning signs that may signal a return to deviant functioning based on each offender's descriptive offence chain, and the generation of plans to deal with potential future problems including training in adaptive coping strategies (e.g., Eccles & Marshall, 1999).

Kia Marama – an example of a treatment programme for child sexual offenders.

Kia Marama is a group treatment programme for men who have sexually offended against children, located within a medium-security 60-bed unit at Rolleston

Prison near Christchurch, New Zealand (the site from which the sample for the empirical sections of this dissertation was drawn from). The unit is self-contained within the prison, and operates as a therapeutic milieu. The treatment programme is based on cognitive-behavioural principles, with a relapse prevention framework. It should be noted that Kia Marama aims to evolve as research advancements continue to be made. For example, rolling or open-ended groups as advocated by Marshall (Marshall et al., 2006), are being trialled at the unit at the time of writing. The programme description outlined here (see Hudson et al., 1998, for a more detailed description) generally refers to the period during which the sample utilised for the empirical section of this work were receiving treatment, from around 1993 to 2000.

In the Kia Marama treatment programme, groups of 10 men meet with a therapist three times per week for roughly 2.5 hour sessions. Treatment lasts for around 33 weeks, including 2-week periods of assessment at the start and end of the programme. Groups progress through eight modules, labelled: Norm Building; Offence Chains; Arousal Reconditioning; Victim Impact and Empathy; Mood Management; Relationship Skills; Sexuality; and Relapse Prevention.

In Norm Building, the rules of group conduct (e.g., confidentiality, participation, communication guidelines) are established collaboratively, and members engage in initial self-disclosure by introducing themselves and discussing their backgrounds and reasons for entering the programme. Offence Chains is a fundamental module, to which the greatest number of sessions is devoted. Each member develops his personal offence chain and presents it to the group. Through several revisions the goal is for the offender to demonstrate comprehension of the phases in his chain and the links between them, as well as a sense of ownership or responsibility for the offending and its build-up. Arousal Reconditioning involves

behavioural interventions to reduce deviant sexual arousal and strengthen arousal to appropriate thoughts and images; specifically covert sensitisation, directed masturbation, and satiation procedures (described above) are implemented at Kia Marama. In the Victim Impact and Empathy module, groups first identify and learn about the immediate, short-term, and long-term effects of sexual abuse on victims in general, and then discuss these with reference to their own victims. Several carefully selected readings and videos depicting victims' experiences are used to facilitate perspective-taking, and again the group are asked to relate their insights to the perceived effects of the victims of their own offending. In the past a guest speaker, a sexual abuse survivor, was invited to address the group about the impact of the abuse in general and on themselves, and facilitate a discussion (this practice was discontinued in 2001). Finally, each offender writes an "autobiography" from the perspective of his own victim, reading it aloud to set the scene for a role-play in which he alternates between the roles of himself and his victim.

The remaining modules focus on equipping offenders with skills to reduce their risk of reoffending in the future, based on various deficits that commonly feature in the offence chains of sexual offenders (for example deficits in emotion-regulation, self-regulation, social competency, intimacy skills, and knowledge about adult sexuality). Mood Management begins with psychoeducation about the cognitive-behavioural model of mood, followed by teaching of how to identify a range of affective states (such as sadness, anxiety, fear, and anger). Strategies for managing these moods are then taught, including physiological relaxation, exercise and a balanced diet, cognitive restructuring of negative thought patterns, and behavioural skills such as developing positive coping strategies, emotionally expressive communication, assertiveness training, and techniques relating to anger management,

conflict resolution, time management, and problem solving. In the Relationship Skills module, the meaning and value of intimacy in adult relationships is discussed, as well as the blocks or fears that may prevent people from achieving it. Each group member identifies their own dysfunctional relationship style, and learns skills for enhancing intimacy and resolving conflict. The aims of the Sexuality module are to increase accurate knowledge of adult sexuality (including sexual dysfunction), reduce anxieties in discussing sexual matters, and engender attitudes conducive to respectful, consenting, and mutually rewarding sexual intimacy with an adult partner. This module makes use of widely-available educational resources such as videos. The final module is Relapse Prevention, in which relevant offence chain concepts are revised, along with coping strategies appropriate to each of the links, with an emphasis on choices and personal agency. Each group member identifies their own high risk situations and internal and external warning signs, and develops a personal life plan to avoid reoffending.

As noted, Kia Marama operates as a stand-alone unit, with a therapeutic community housing the treatment participants on-site. After completing the programme, it is practice for men to remain within the unit until their release, attending weekly “Graduates” group meeting. This period allows for treatment gains to be consolidated prior to release. Parole conditions for men who have completed treatment at Kia Marama also typically include attendance at monthly follow-up group meetings, held at various locations around the country, in which treatment concepts are revised and guidance can be sought if problems arise.

Does Treatment for Sex Offenders Work?

The primary goal of sex offender treatment is to eliminate or reduce reoffending, reflected in the Kia Marama axiom of “no more victims.” Evidence of reduced recidivism also provides justification for rehabilitation as a major endeavour of justice systems, as opposed to simply punishing the criminal or preventing offending during the period of incarceration only (incapacitation). Thus, programme evaluation is as important as programme implementation, and indeed can guide further theorising and advancements in the field. In this section, research on the effectiveness of treatment programmes at reducing reoffending will be reviewed.

Many authors have pointed out that evaluating treatment effectiveness in this area is a difficult task (e.g., Marques, 1999; Craig, Browne, & Stringer, 2003b; Lösel & Schmucker, 2005). The selection of appropriate control groups is difficult for ethical reasons – the ideal option for methodological rigour would be a randomly assigned untreated group, but withholding treatment poses a considerable risk to society and potential victims (although such a control group was included in a Californian study discussed below, justified by arguments that neither group would receive treatment if not for the trial; Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005). Another problem is the uncertainty of the outcome data. Recidivism base rates are typically low, and a large portion of sexual offences are thought to remain unreported, decreasing statistical power (Hanson, 1997b). There is also potential for very long-term reoffending among sex offenders (Hanson, Steffy, & Gauthier, 1993), so that by the time sufficient outcome data is available the intervention being evaluated could be relatively obsolete compared to new innovations, as well as the problem of eventual reoffenders being erroneously classified as non-reoffenders. Methodological rigour would also require evaluation studies to deliver treatment according to standardised protocols to minimise effects

beyond those attributable to the programme, such as therapist effects. However as discussed above, recent research on the importance of certain therapist characteristics (including flexibility) for successful treatment suggests that over-manualised approaches may reduce effectiveness (Marshall, 2005). Because of these and other difficulties, the interpretation of treatment effectiveness research has been controversial, and debate on this question continues.

Meta-analyses have at least been able to redress the base rate and statistical power problem, by amalgamating the results of many outcome studies to produce overall effect sizes for treatment with very large sample sizes. Hall (1995) conducted a meta-analysis on 12 treatment evaluation studies published between 1988 and 1994, when promising comprehensive (multi-component) programmes were beginning to have acceptable follow-up periods. Overall, a small but significant treatment effect size of $r = .12$ was found, with the amalgamated treatment groups sexually reoffending at a rate of 19%, compared to 27% of the comparison groups (untreated or alternative treatment, e.g., non-cognitive-behavioural group therapy). Hall also found that cognitive-behavioural treatments were significantly more effective than other types of therapy (similar effect sizes were found for hormonal therapy, but practical disadvantages were discussed). A larger and more recent meta-analysis initiated by the American Association for the Treatment of Sexual Abusers (ATSA Collaborative Study, Hanson et al., 2002) included 43 studies, with a combined sample size of over 9,000. Overall, treated offenders had a lower sexual reoffence rate (12%) than comparison groups (who received no treatment or treatment judged to be inadequate or inappropriate; 17%), again a small but significant effect (odds ratio = 0.81). When their analyses were restricted to studies employing contemporary treatment approaches (cognitive-behavioural for adults or systemic for adolescents)

and randomly or incidentally assigned control groups (as opposed to groups confounded by treatment refusers) the effect was larger (odds ratio = 0.60), corresponding to a sexual recidivism rate of 10% for treated offenders and 17% for the comparison groups. The effectiveness of treatment and the particular robustness of cognitive-behavioural approaches were also supported in Lösel and Schmucker's (2005) cross-cultural meta-analysis.

These results are promising, however Rice and Harris (2003) argued that the majority of the studies analysed by Hanson et al. (2002) used methodologies too weak to address the question of treatment effectiveness, such as the use of comparison groups that were not randomly assigned and analyses confounded by treatment refusers and dropouts. They reiterated that no treatment effect was found among the small number of randomised control trials included in the meta-analysis. Additionally, an influential and highly-anticipated randomised control trial of sex offender treatment conducted in California (Marques et al., 2005) also failed to find a significant treatment effect when comparing the survival rates of those who received treatment and volunteer and non-volunteer controls (more positive earlier data from this study were included in the meta-analyses by Hall, 1995, and Hanson et al., 2002). Although this study could be viewed as simply not supporting the efficacy of sex offender treatment, Marques et al. (2005) discussed at length other possible reasons for the failure of their cognitive-behavioural relapse prevention programme. These included the hospitalisation of the treatment group, overall higher risk level of the treatment group, lack of external motivation protocols, low recidivism base rates overall, a possible mismatch between risk level and treatment intensity, limited scope of the original relapse prevention model utilised (although it was state of the art at the time of programme inception), and over-manualisation of the treatment and aftercare

components. Additionally, some within-group differences suggesting the efficacy of treatment for some participants were found among those who received treatment; these will be discussed in a later section of this review.

Therefore, although this issue continues to generate debate within the literature, the balance of evidence appears to show that treatment can be effective at reducing the recidivism of sex offenders, particularly when based on contemporary application of cognitive-behavioural principles. The effectiveness of the Kia Marama programme in particular was examined by Bakker, Hudson, Wales, and Riley (1998), who reported a significant difference between the sexual reconviction rates of the first 238 men to complete the programme (10%) with an incidental matched control group (22%). The control group was made up of men who were imprisoned for similar offences before the programme was available (time at large was taken into account in these analyses). A further effectiveness investigation was conducted following improvements to the Kia Marama programme, including the introduction of bicultural treatment concepts, establishment of the therapeutic community in the unit, and the application of more general improvements to the CBT model for sex offenders (Rutherford & Grace, 2004). In these analyses, reoffence rates were found to be significantly lower for those who completed treatment after 1994 (taking risk and time at large into account), with an overall 4% reconvicted across the follow-up period of up to seven years.

Considerations of the benefits of sexual offender treatment, however, go beyond statistical significance. Marshall and McGuire (2003) compared the averaged effect sizes of sex offender treatment based on four meta-analytic reports (including the two discussed above: Hall, 1995; and Hanson et al., 2002) to effect sizes for treatment of general offending (non-sexual), mental health and psychological

interventions, and interventions for physical health problems. Effect sizes are used to interpret the magnitude of a treatment effect independently of sample size, and can therefore inform as to whether a statistically significant difference is of practical concern. They concluded that the effect sizes for sex offender treatment (e.g., 0.28 for current treatments in the ATSA collaborative study; Hanson et al., 2002) were comparable to those shown for interventions considered efficacious for various other problems, and were in fact far superior to effect sizes that have dramatically influenced policy in regard to medical problems, such as daily aspirin for the prevention of heart attacks (effect size of 0.03). However, it is important to note that these calculations are based on the same (predominantly incidentally assigned) control groups criticised by Rice and Harris (2003) as being inappropriate.

Nonetheless, there are important social and economic benefits to be considered of even small reductions in recidivism. Donato and Shanahan (2001) conducted an analysis of the economical costs and benefits associated with sex offender treatment programmes. They highlighted several areas of savings as a result of reduced recidivism, including formal costs to society resulting from court and incarceration costs, and expenditures on police and social welfare services, as well as the more intangible costs associated with the sequelae of child sexual abuse, such as physical injury and illness, psychological and emotional suffering, psychiatric disorders (e.g., trauma, depression, and anxiety), and the potential for inter-generational abuse (i.e., victims themselves becoming perpetrators). According to Donato and Shanahan's (2001) analyses, a reduction in recidivism of 12 percentage points (as was reported in the Kia Marama analysis by Bakker et al., 1998) would result in an expected net benefit of \$120,000 per group of 10 treated offenders based on tangible savings, plus up to an additional \$106,800 per group associated with more

intangible costs (Donato & Shanahan's figures are based on 1990 U.S. dollars and estimate the costs of treatment for one offender at \$10,000, somewhat higher than the Kia Marama estimate of \$NZ8,403 reported by Bakker et al. in 1998). Even if the reduction in recidivism was halved (to six percentage points), the state would still yield a net saving per treated offender even without considering the intangible costs of recidivism, according to their figures. Of course, as discussed by Marshall and McGuire (2003) and Donato and Shanahan (2001), recidivism analyses typically examine recidivism as a dichotomous variable (i.e., yes or no), thereby not taking into account that many offenders have multiple victims. Incorporating this into considerations, the social and economic benefits of reductions in recidivism increase exponentially.

Overall, despite the difficulties with evaluative research in this area and the lack of a strong consensus to date, the considerations presented above point to the fair conclusion that rehabilitative efforts for sex offenders such as the Kia Marama programme are a worthwhile and socially-responsible endeavour of the criminal justice system. Ongoing advancements in this field may contribute to improving treatment effectiveness and reducing recidivism even further. One important new direction, given the heterogeneity of sex offenders, would involve evaluating sex offenders differentially, thereby addressing issues such as "what works for which types of offenders?," rather than simply "is treatment effective?" (Marques, 1999; Lösel & Shmucker, 2005). As noted by Bickley and Beech (2003), evaluating treatment effectiveness among sex offenders as a homogeneous group has the potential to mask significant treatment effects among certain types of offenders, a factor that has perhaps contributed to the controversy in this area of research.

As important as it is, treatment is not the only endeavour of researchers and clinicians in the criminal justice field. The next section of this review focuses on techniques and challenges involved with the practice of risk assessment, particularly among sex offenders who have received treatment. The research outlined above has shown that treatment can reduce risk of reoffending; it is important that this be incorporated into professional estimations of risk in a valid way.

Risk Assessment Among Sex Offenders

Assessing risk among sex offenders is an important task for clinicians working in the criminal justice field. In most cases, estimates of an offender's likelihood of reoffending are used to inform decisions with real-life consequences, such as those pertaining to sentencing, institutional security ratings, provision of treatment, release planning, and parole conditions. It is therefore important that risk estimates show external validity in terms of predictive accuracy, so that an appropriate balance can be struck between the rights of offenders and the protection of society. Overestimating risk (i.e., a high false-positive rate) in some jurisdictions can result in unjustified extended periods of detainment or over-intensive treatment or supervision. Aside from jeopardising the rights of the offender, this would be in contrast to the risk principle, and a waste of resources. Conversely, underestimating risk (a high false-negative rate) may impact negatively on the safety of society by failing to provide adequate treatment or supervision to high-risk individuals.

First Generation of Offender Risk Assessment: Unstructured Clinical Judgement

Andrews and Bonta (2003) have described the development of assessment approaches among general offenders as having evolved across three "generations," and the process for sexual recidivism has mirrored this. The first generation refers to subjective and unstructured clinical judgement, in which the professional forms an opinion of whether or not an individual is likely to reoffend based on their training and experience. This approach has now been discredited due to research indicating poor inter-rater reliability and poor predictive accuracy (Grove & Meehl, 1996;

Hanson & Bussière, 1998). In particular, unstructured clinical judgement can be subject to an overestimation bias in which consideration of base rates is neglected (Craig, Browne, Stringer, & Beech, 2004). Although accurate recidivism rates among sex offenders are unfortunately unknown (Prentky, Lee, Knight, & Cerce, 1997), research has indicated that it is by no means inevitable that someone convicted for a sexual offence will reoffend once released from prison. In a meta-analysis of 82 studies with an average follow-up period of five to six years, the observed sexual recidivism rate was 13.7% (Hanson & Morton-Bourgon, 2005). Even with a follow-up period of 25 years, and using the more inclusive outcome of charges (rather than convictions), Prentky, et al. (1997) reported sexual recidivism rates of 26% for rapists and 32% for child molesters. These low base rates make accurate prediction of recidivism (traditionally the goal of offender risk assessments) difficult.

Risk Factors – Static and Dynamic

Despite the low base rates, research has shown that certain sexual offenders have a much higher rate of recidivism than others (G. T. Harris et al., 2003). Variables that have been shown to be linked to increased rates of recidivism are known as risk factors, and can be divided into two major types: static and dynamic. Static risk factors are those that are fixed or unchangeable, are often historical in nature, and are markers for long-term propensities towards sexual offending (Hanson, 1998). Examples of static factors include those relating to offending history and demographic variables. Dynamic factors are those that are related to recidivism, have the potential to change (i.e., through treatment), and should be associated with increased or decreased recidivism when changed (Hanson, 1998). Because of this, dynamic factors

are usually the targets for change in sex offender treatment programmes; examples include variables relating to attitudes, social and emotional functioning, and sexual interests (Craissati & Beech, 2003). Although changeable, most dynamic factors that are treatment targets are relatively enduring; these are referred to as stable dynamic factors. By contrast, acute dynamic factors are those that change more rapidly and are more closely related to the timing of recidivism (Hanson, 1998), such as substance intoxication or emotional collapse.

This traditional distinction between static, stable dynamic, and acute dynamic risk factors has been redefined in a series of articles by Beech and Ward (Beech & Ward, 2004; Ward & Beech, 2004), in an attempt to provide some integration between the areas of risk and etiology. Their model of “The Etiology of Risk” posits so-called stable dynamic risk factors as dispositional (i.e., trait) vulnerability factors for sexual offending, arising etiologically out of various developmental adversities (including abuse, rejection, and attachment problems). Traditionally-defined static factors are considered in the model to be historical markers for these trait vulnerabilities. Finally, acute dynamic factors are hypothesised to be the state expressions of these vulnerabilities in response to contextual precipitants (such as victim access, relationship conflict, and substance abuse).

Second Generation of Offender Risk Assessments: Actuarial Risk Instruments

The second generation of general offender risk assessment (as described by Andrews & Bonta, 2003) began to develop in the 1970s, as the limitations of unstructured clinical judgement became clearer and knowledge of empirically-validated risk factors was increasing. In this approach, objectivity and empirical

validity were emphasised by the development of structured actuarial instruments made up of risk items previously found to be related to recidivism. Typically, these instruments involve each risk item being rated according to standardised scoring protocols, with total scores (usually derived additively) corresponding to empirically derived risk estimates. Reflecting the second generation trend towards increased objectivity, actuarial measures predominantly consisted of static risk factors initially, favoured due to their superior objectivity and inter-rater reliability compared to dynamic factors, as well as the paucity of research into dynamic factors at the time.

Although actuarial scales for general offending have been in use for some time (e.g., Statistical Information on Recidivism Scale, SIR; Nuttfield, 1982, cited in Andrews & Bonta, 2003), the applicability of such measures for assessing the risk of sexual recidivism is questionable (although many have been shown to predict general recidivism among sex offenders), as many contributing factors are unique to sexual offending, such as sexual deviance (Hanson & Bussière, 1998). Subsequently, the mid-1990s saw a surge of research into risk factors for sexual recidivism in particular. Hanson and Bussière (1998) conducted a meta-analysis of predictors of sex offender recidivism from 61 studies. The majority of the risk factors that had been included in studies by that stage were either static or extremely stable. Factors identified by Hanson and Bussière that had a weighted average correlation (after correcting for differences in base rates) with sexual recidivism of at least $r = .10$ and were replicated across at least four studies included the following: Demographic factors – young age; single marital status; General criminality – antisocial personality disorder; total number of prior offences; Sexual criminal history – prior sexual offences; stranger victims; extrafamilial victims; male child victims; early onset of sex offending; diverse history of sexual crimes; Sexual deviancy – phallometric assessment

indicating interest in children; any assessment of deviant sexual preference; and Clinical presentation – failure to complete treatment. The largest predictor variables were those related to sexual deviancy (Hanson & Bussière, 1998).

These data were used by Hanson (1997a) to develop a second generation instrument specifically for the purpose of assessing risk of sexual recidivism, called the Rapid Risk Assessment for Sex Offender Recidivism (RRASOR). The RRASOR is a simplistic scale that includes four items pertaining to age, prior sex offences, and victim type. Other well-known second generation actuarial risk measures for sexual offenders based on empirically validated risk factors include the Static-99 (Hanson & Thornton, 1999), and the Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998). The Static-99 is a 10-item solely static scale that is a combination of the RRASOR and the Structured Anchored Clinical Judgement – Minimum version (SACJ-Min; Hanson & Thornton, 1999) – a staged risk measure based on sexual and non-sexual criminal history, as well as potentially aggravating factors pertaining to victim type, marital history, and history of non-contact sex offences. The SORAG was based on the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993), and was designed to assess violent recidivism (including sexual violence) among sex offenders. It includes 14 items pertaining to historical social and developmental factors and alcohol abuse, personality disorders and schizophrenia, sexual and non-sexual criminal history and past failure on conditional release, phallometrically assessed deviant sexual interests, and age. An important item in the SORAG is the individual's score on the Psychopathy Checklist-Revised (PCL-R; Hare, 1991), a widely-used reliable and valid measure of the personality construct of psychopathy which has been shown to be robustly related to recidivism, particularly violence (Hemphill, Hare & Wong, 1998). Psychopathic traits

measured by the PCL-R include manipulativeness, superficiality, and emotional shallowness (Hare, 1991).

The comparative predictive validity of these second generation risk instruments for sexual recidivism was evaluated by Barbaree, Seto, Langton, and Peacock, (2001), and G. T. Harris et al. (2003). Both studies reported their results in terms of the area under the receiver operating characteristic curve (AUC), a preferable statistic compared to correlations which are influenced by the problem of low base rates. AUC figures represent the probability that a randomly selected recidivist would have a higher score on the prediction measure than a randomly-selected non-recidivist; scores range from 0 to 1, with an AUC of .5 representing prediction at the chance level and higher scores representing better predictive validity. Both studies reported significant ($p < .05$) AUC values in the range of .59 to .77 for the prediction of sexual recidivism using the SORAG, RRASOR, and Static-99, with differences in the predictive validity of the measures being insignificant.

Given the multi-determined nature of recidivism, this level of predictive validity reflects a substantial improvement relative to unstructured clinical judgement. However, there are some obvious limitations to static actuarial methods despite the superior predictive validity. A focus on static predictors results in many theoretically relevant dynamic factors being neglected, despite the increasing body of research into their empirical validity (e.g., Hanson & Harris, 2000; Hudson, Wales, Bakker, & Ward, 2002). Static factors, being historic and unchangeable by definition, offer no indication of appropriate targets for interventions aimed at reducing an offender's recidivism risk. Likewise, measures of static risk are not sensitive to change, and so an offender's risk estimation, based on the Static-99 for example, would remain exactly the same even after successful participation in treatment. This is problematic,

given the evidence (outlined in the previous section of this review) that treatment can be effective at reducing recidivism. Subsequently, there was an increased research interest in dynamic risk factors for sexual recidivism, and how these could best be incorporated into risk assessments, providing the impetus for the third generation of risk assessment (Andrews & Bonta, 2003).

Third Generation Risk-Needs Assessments: Incorporation of Dynamic Factors

Andrews and Bonta (2003) define third generation risk assessments as being distinguished from second generation instruments by the systematic and objective measure of offender needs. The benefits of this approach are twofold: incorporating dynamic factors into risk assessments may improve the predictive accuracy of static factors alone; and knowledge of an offender's dynamic risk factors or criminogenic needs can inform as to appropriate treatment targets to reduce their risk.

Seven years after Hanson and Bussière's (1998) meta-analysis was published, Hanson and Morton-Bourgon (2005) reported the results of an updated meta-analysis of risk factors for sexual recidivism which aimed to incorporate newer research, particularly studies that looked at dynamic factors. This review had an amalgamated sample size of 29,450 sex offenders from 82 studies. Of these, 35 had been previously reviewed by Hanson and Bussière, 10 contained updated information, and 37 were new. Their major finding was the identification of deviant sexual preferences and antisocial orientation as the major predictors of sexual recidivism. These two dimensions in fact underlie many of the static risk markers identified in Hanson and Bussière's earlier meta-analysis and are incorporated within actuarial risk instruments; for example, male victims, non-contact sex offences, and number of prior

convictions (all from the Static-99; Hanson & Thornton, 1999). Hanson and Morton-Bourgon (2005) also found small but significant predictive effects for the general categories of sexual attitudes (particularly tolerant attitudes towards sex crime) and intimacy deficits (particularly emotional identification with children, and conflicts in intimate relationships). These findings were broadly reiterated in a qualitative review of dynamic risk factors (Craissati & Beech, 2003), in which the authors concluded that there appeared to be a convergent view regarding the importance of pro-offending attitudes, intimacy deficits, deviant sexual interests, and general self-regulation problems as dynamic risk factors capable of predicting sexual recidivism independently of static factors. Hanson and Morton-Bourgon (2005) also identified several areas of dynamic risk that could theoretically be conducive to change in treatment, and were reliably empirically related to sexual recidivism. These were: any deviant sexual interest; sexual preoccupations; antisocial personality disorder; Psychopathy Checklist – Revised; general self-regulation problems; employment instability; and hostility.

Incorporating some consideration of empirically validated dynamic factors into risk assessments should theoretically improve predictive accuracy, given the limitations of purely static instruments discussed above. However, it is important that the methods employed to do this are structured and empirically validated. Subjectively “adjusting” an actuarial risk estimate by considering dynamic risk factors would effectively re-introduce a degree of clinical judgement (shown to have minimal predictive value), potentially diluting, rather than enhancing, predictive accuracy (Hanson, 1998). Examples of sex offender risk instruments that adopt a relatively unstructured or “guided clinical” approach to considering dynamic variables include the Sexual Violence Risk-20 (SVR-20; Boer, Wilson, Gauthier, & Hart, 1997, cited in

Craig, Browne, & Stringer, 2003a) and the Multifactorial Assessment of Sex Offender Risk for Recidivism (MASORR; Barbaree et al., 2001), both developed prior to the widespread availability of validated actuarial scales. The MASORR includes consideration of the degree of change achieved in treatment, motivation, and clinical impression, subjectively combined with the pre-treatment estimate. Barbaree et al. (2001) found that the MASORR was not significantly predictive of sexual recidivism either pre-treatment or post-treatment (AUC values were .61 and .60 respectively).

Several authors have since developed protocols for the objective assessment of dynamic risk among sex offenders, and evaluated their empirical validity for predicting recidivism alongside static estimates. Hanson and Harris (2001) developed an instrument to assess stable and acute dynamic risk among sex offenders called the Sex Offender Need Assessment Rating (SONAR), based on data obtained from file reviews and interviews with community supervision officers of recidivist and non-recidivist sex offenders. This instrument has since been developed further, and divided into two scales called the STABLE 2007 and the ACUTE 2007, designed to be used in conjunction with the Static-99 as a hierarchical risk assessment protocol for the initial evaluation and long-term supervision of sex offenders (Hanson, Harris, Scott, & Helmus, 2007). The STABLE 2007 consists of 13 items within five domains of dynamic risk factors for sex offender recidivism: social influences; intimacy deficits; general self-regulation; sexual self-regulation; and cooperation with supervision. Empirically-guided protocols for combining Static-99 and STABLE 2007 scores to form an overall initial risk estimate are provided. For example, an offender with a Medium-Low score on the Static-99 and a High score on the STABLE 2007 would qualify for an overall risk rating of Moderate-High (Hanson et al., 2007). The ACUTE 2007 was designed for more frequent use by supervision officers in the

community (at one-month intervals) to assess for changes that may signal an increased or decreased risk of recidivism, and protocols for combining this with the initial risk estimate to determine supervision priority are provided. Acute items relating to sexual risk include victim access, hostility, sexual pre-occupation, and rejection of supervision.

Recent research (Hanson et al., 2007) supported the predictive validity of the risk framework. The STABLE 2007 had a significant AUC value of .67, and scores were significantly predictive of sexual recidivism after controlling for the Static-99. The combined risk estimate based on Static-99 and STABLE 2007 scores had an AUC of .76. Change on the STABLE 2007 was analysed by re-administering the measure after six months in the community. Test-retest scores were found to be highly consistent (ICC = .79), and although offenders on average were assessed as having significantly lower scores on the re-test, the amount of change was unrelated to recidivism. The ACUTE 2007 risk items for sexual or violence risk provided significant incremental validity in predicting sexual recidivism after controlling for the Static-99 and the combined risk estimate based on the Static-99 and STABLE 2007.

Beech (1998) used cluster analysis to develop protocols for the psychometric assessment of dynamic factors (pro-offending attitudes and socio-affective problems) among child molesters as part of the Sex Offender Treatment Evaluation Project (STEP) in the U.K. According to these protocols, offenders can be classified as either High Deviance or Low Deviance on the basis of the deviation of their scores from non-offender norms on a battery of relevant psychometric tests. Relative to non-offenders, High Deviance men have higher levels of cognitive distortions about children, distorted attitudes about their victims, emotional identification with children,

sexual obsessions, and self-reported patterns of sexual deviance. High Deviance offenders also show various socio-affective difficulties including difficulty forming intimate attachments with adults, low self-esteem, and under-assertiveness. Low Deviance offenders show a lower level of social adequacy problems, and do not have generalised cognitive distortions about children or emotional identification. However, both groups were found to have poor empathy for their own victims (Fisher, Beech, & Browne, 1999).

The relevance of Beech's (1998) typology to sexual recidivism risk was supported in a six-year follow-up study of child sexual offenders who received community-based treatment (Beech, Friendship, Erikson, & Hanson, 2002). The Deviancy classification was significantly related to sexual reconviction rates, with 30% of the High Deviancy group reoffending compared to 3% of Low Deviancy offenders (odds ratio of 12.7). Logistic regression analyses additionally showed that the Deviancy classification contributed to reconviction independently of static risk as measured by the Static-99. Beech et al. (2002) advised that their results should be interpreted cautiously given the small sample size of 53, and that their approach did not represent an applied risk assessment methodology given the dependence on local sample means. Like Hanson et al.'s (2007) analysis of change on the STABLE 2007, Beech et al. (2002) also pointed to the high test-retest stability (typically around .80) of the measures used to assess supposedly dynamic or changeable variables, suggesting the relatively enduring nature of the characteristics. Given that pre-treatment assessments were able to predict recidivism up to six years later, further research would be required to determine whether reductions in Deviancy would be associated with decreased recidivism (Beech et al., 2002). Notably, these empirical findings regarding the durable nature of stable dynamic risk appear consistent with

Beech and Ward's conceptualisation of these factors as enduring psychological or dispositional characteristics, akin to personality traits (Beech & Ward, 2004; Ward & Beech, 2004).

Also in the U.K., Thornton (2002) developed a comprehensive framework for sex offender risk assessment called Structured Risk Assessment (SRA) or Structured Assessment of Risk and Needs (SARN). This scheme involves four parts: an assessment of static risk (i.e., Static-99); an Initial Deviance Assessment (IDA) in which four domains of dynamic risk are assessed; an Evaluation of Progress (based on response to treatment); and Risk Management (based on offence-specialisation and acute risk factors). Within the IDA, "deviance" is defined as the extent to which an offender's functioning is dominated by psychological factors that contribute to his offending. The four identified dynamic risk domains are sexual interests, distorted attitudes, socioaffective functioning, and self-management. These are assessed psychometrically, with the exception of sexual interests, which may be assessed using methods such as phallometry. An offender is classified as dysfunctional on a domain if his scores are above the sample mean on at least two of the psychometric measures from that domain. An overall Deviance classification (IDA) is then given, with High Deviance meaning the offender was classified as dysfunctional in at least 2 domains, Moderate for one domain, and Low for none. Thornton (2002) evaluated the predictive validity of the IDA. It was found that sexual reconviction was monotonically associated with overall deviance classification: 0% of Low Deviance offenders, 5% of Moderate Deviance; and 15% of High Deviance recidivated. The Deviance classification had an AUC of .78, and was independently predictive of recidivism after controlling for the Static-99.

Recently, a framework for assessing dynamic risk among child molesters has been developed in New Zealand, using an individual differences approach with data obtained from the battery of psychometric tests administered for 492 child molesters who had completed treatment at Kia Marama (Allan, Grace, Rutherford, & Hudson, 2007). Exploratory factor analysis of the pre-treatment psychometric data revealed four primary dimensions – social inadequacy, sexual interests, anger/hostility, and pro-offending attitudes. Similarities between this model and others were noted, particularly Thornton's (2002) IDA domains. To analyse the relationship between these factors and recidivism, Allan et al. (2007) determined factor scores for each offender, calculated as the average of the standardised scores for the tests loading onto each factor. All four factors were found to be significantly predictive of recidivism; sexual interests and pro-offending attitudes (respective AUCs of .72 and .70) to a higher degree than social inadequacy and anger/hostility (AUCs of .62 and .60). An Overall Deviance score was then determined based on the number of domains for which the offender scored above the sample mean, with sexual interests and pro-offending attitudes being more heavily weighted to reflect their stronger association with recidivism. Overall Deviance scores had an AUC of .76 with regard to sexual recidivism, and were significantly predictive after controlling for the Static-99. An Overall Risk score (the equally weighted average of Static-99 and Deviance scores) had an AUC of .81.

Fourth Generation? New Directions in Risk Assessment

A criticism of both the second and third generations of risk assessment is that, although their evidence-base represents an improvement over clinical judgement, they

often lack a theoretical foundation, the items being solely chosen due to their statistical relationship with recidivism (Andrews & Bonta, 2003). Additionally, although the examples of dynamic risk frameworks discussed above (Allan et al., 2007; Beech, 1998; Thornton, 2002) show sound predictive validity, the incorporation of psychometric batteries make them resource-intensive, and therefore not easily employed in other settings. The protocols for assessing change across treatment (if included) tend to be underdeveloped, likely due to research into the assessment of change across treatment still being in an early stage. As noted by Beech et al. (2002), it is likely that the dynamic risk factors assessed within these frameworks are relatively enduring, given that pre-treatment assessments were predictive of recidivism after several years follow-up despite receiving treatment aiming to address these need domains (this was also the case in Allan et al., 2007). It is therefore unclear how useful these third generation risk frameworks might be for the purpose of incorporating reductions in dynamic risk as a result of gains made in treatment.

A comprehensive discussion of studies that have employed some form of structured assessment relating to proximal outcomes of treatment will form the next section of this review. One study included in that section describes the development and validation of a new risk instrument, the Violence Risk Scale: Sexual Offender version (VRS:SO; Olver, Wong, Nicholaichuk, & Gordon, 2007; based on the Violence Risk Scale, VRS; Wong & Gordon, 2006), which the authors described as a dynamic actuarial instrument, and suggested may represent a new, “fourth generation” of risk assessment (Olver, 2003). The VRS:SO includes both static and dynamic risk components, as well as protocols for the structured and objective measurement of change across treatment, with theoretical grounding in the Transtheoretical Change Model (Prochaska, DiClemente, & Norcross, 1992). The dynamic items were selected

based on their theoretical and empirical relevance to recidivism, and comprise three domains: sexual deviance, criminality, and treatment responsivity. Pre-treatment scores on the dynamic items inform as to the criminogenic needs and potential treatment targets for an offender, as well as their motivational stage for each (pre-contemplation, contemplation, preparation, action, or maintenance). Post-treatment scores are obtained by adjusting the pre-treatment ratings on the dynamic items according to progression through the motivational stages.

In the initial validation study (Olver et al., 2007), the VRS:SO authors reported good inter-rater reliability (significant intraclass correlation coefficients for the dynamic scale in the range of .74 to .79), and predictive validity with regard to sexual recidivism. AUC values for pre-treatment and post-treatment scores were .71 and .72 respectively (findings relating to the assessment of change across treatment with the VRS:SO will be reviewed in the next section). Aside from the predictive validity, the benefits of the VRS:SO include the valuable treatment information relating to identification of criminogenic needs, the empirical and theoretical underpinnings of the measure, and the simplicity of having the assessment of static and dynamic risk, responsivity issues, and treatment gain all incorporated into a single instrument. The VRS:SO therefore appears very promising as a multi-purpose assessment tool for sex offenders, however it is important that the validity of the instrument is independently evaluated in different settings.

Another potential future direction in sex offender risk assessments has been highlighted by Doren (2004), who noted that all of the risk assessment approaches reviewed above conceptualise risk as varying linearly across a single dimension from low to high. Even when multiple domains are considered, risk ratings tend to be averaged or otherwise reduced down to a unidimensional estimate, with all identified

risk factors conceptualised as increasing or decreasing risk along that single continuum. Doren (2004) argued that this may be too simplistic a view, and discussed research findings indicating support for a multidimensional model for sexual recidivism risk. In particular, he noted the robust indication from independent studies carried out globally and employing several different methodologies (including meta-analyses, factor analyses, pathway analyses, and qualitative reviews), that there appeared to be two major risk dimensions that are both strongly related to risk but unrelated to each other – sexual deviancy, and psychopathy/general criminality. Although there is evidence for an interactive effect, whereby the combined presence of both risk factors is associated with a further increase in risk (e.g., G. T. Harris et al., 2003), both dimensions appear to be independently predictive of recidivism. Doren (2004) discussed the clinical implications of a multidimensional model for treatment and risk assessment, specifically the importance of an evaluator considering each dimension, with evidence of risk from any single dimension being indicative of overall risk (analogous to a doctor carrying out several different necessary blood tests, with problems diagnosed from any of these being indicative of overall health problems). Doren's ideas point to the importance of identifying specific clusters of risk factors associated with the specific offence pathways of different offenders. The suggestion of possible interactive effects between risk variables is also intriguing and warrants further exploration, perhaps particularly with regard to psychopathy, which has been identified not only as a risk factor among sex offenders, but also a responsivity factor that may impact response to treatment (Rice, Harris, & Cormier, 1992). This potential interaction effect has been investigated in several studies (e.g., Seto & Barbaree, 1999), and will be discussed further in the next section.

As noted, the next section of this review consists of an overview of studies that have employed some form of structured assessment relating to gains made in treatment. Clinical and research rationales for assessing proximal treatment outcome will be outlined, followed by the results of a comprehensive literature search of previous conceptualisations and operationalisations, with a focus on the relationship between treatment outcome and recidivism.

Treatment Outcome – Conceptualisations and Relationship to Recidivism

In this section, previous research that has considered and attempted to measure the concept of treatment outcome among sex offenders will be reviewed. “Treatment outcome” in this dissertation refers to the degree of proximal change, progress, or gains achieved by offenders as a result of treatment (i.e., as measured immediately following treatment), as opposed to longer-term conceptualisations of outcome such as recidivism. Put simply, considering the treatment outcome of a particular offender involves asking the question: “How well has he done in treatment”?

Why Measure Treatment Outcome?

Assessing the treatment outcome of individual offenders has several important applications, both clinical and research. Clinically, information on gains made in treatment can be a pertinent consideration for post-treatment risk assessments (as discussed in the previous section). As Hanson, Cox, and Woszczyzna (1991) argued, “if the correctional and justice systems responses to sexual offenders are influenced by whether the offenders have benefitted from treatment, then it becomes important to be able to assess the effectiveness of treatment for particular offenders” (p. 179). As will be seen, however, treatment outcome is a separate (though related) concept to post-treatment dynamic risk, and as such its applications go beyond post-treatment risk assessment.

In terms of research applications, taking response to treatment into account may help to clarify the controversy with regard to the efficacy of sex offender treatment programmes discussed previously in this review (Scalora & Garbin, 2003).

The typical methodology employed in evaluation studies is to compare the reoffence rates of treated versus untreated samples. However, compared to other types of treatment populations for which this design may be appropriate (e.g., relapse rates of people treated for depression or anxiety disorders), individuals undergoing treatment for sexual offending may be different in that their levels of intrinsic motivation to change could be lower and more variable. Treatment could be court-mandated, or carry secondary motivating factors such as favourable parole board reports and earlier release. Evaluating all treatment completers as a single cohort may therefore mask or wash out substantial treatment effects among those who do make positive use of treatment (Anderson, Gibeau, & D'Amora, 1995; Bickley & Beech, 2003). Further to this, Seager, Jellicoe and Dhaliwal (2004) have suggested that comparing the recidivism rates of those who appear to have benefited from treatment to those who have not may be a more appropriate way of evaluating the effectiveness of a programme than comparing treated versus untreated samples, to avoid the confounding effects of those who either refuse or drop out of treatment.

Marques (1999) discussed the value in researching the relationship between specific components of treatment and recidivism – for example, to explore whether change on targeted dimensions is related to risk of reoffence. Targeting the components most associated with recidivism could result in ongoing improvements in recidivism rates among treated offenders. Research linking specific in-treatment gains to recidivism could also address the important question of whether so-called dynamic risk factors are in fact changeable, and whether such changes are linked to reduced recidivism. Finally, a further potential research application of treatment outcome measures could involve the exploration of responsivity issues. Given that that the principle of specific responsivity is concerned with maximising individuals' capacity

to benefit from treatment, proximal treatment outcome would be an appropriate dependent variable for studies exploring the relationship between various offender factors and treatment gains.

For all of the clinical and research applications of treatment outcome assessments outlined above, the importance of outcome measures being associated with recidivism is clear, since that is the ultimate outcome of sex offender treatment. For example, in terms of using evaluations of treatment outcome clinically to inform post-treatment risk assessments, it would be necessary that such assessments actually had predictive validity with regards to reoffending. Likewise if such ratings were to be applied in research as alternate means of testing the efficacy of a programme or programme components, or as an outcome measure to explore responsivity issues. Below, previous conceptualisations of treatment outcome and attempts to measure the concept will be reviewed, with a particular focus on the relationship between these various methods and sexual recidivism.

Conceptualisations of Treatment Outcome – Literature Search

Despite the many important clinical and research applications, relatively little research has been done on methods to evaluate proximal treatment outcome among sex offenders in systematic, reliable, and valid ways, and even fewer studies have examined the relationship between treatment outcome and recidivism rates. A comprehensive search for published studies incorporating assessments of treatment outcome among sex offenders (described in more detail below) identified only two with the direct primary goal of developing a systematic, reliable, and valid method of assessing clinical progress on multiple components among treated sex offenders:

Standard Goal Attainment Scaling for Sex Offenders (SGAS; Hogue, 1994); and the Sex Offender Treatment Rating Scale (SOTRS; Anderson et al., 1995). Although developed over a decade ago, each of these measures has been applied to assess treatment outcome in only a few subsequent published studies (e.g., Stirpe, Wilson & Long, 2001; Levenson & Macgowan, 2004).

The search of the literature also yielded several studies in which some assessment of short-term treatment outcome was carried out in order to address other primary research questions, such as overall treatment efficacy (e.g., Quinsey, Khanna, & Malcolm, 1998), risk assessment (e.g., Olver et al., 2007), factors affecting treatment outcome (e.g., Marshall et al., 2002), and interactions between treatment outcome and other factors such as psychopathy on recidivism (e.g., Seto & Barbaree, 1999).

Table 1 presents all of the relevant published studies found by conducting an extensive search of the literature on the treatment of adult sexual offenders. This search involved the first step of entering various keywords into the PSYCINFO database, such as “treatment outcome,” “treatment progress,” “treatment change,” “treatment success,” “treatment behavior/behaviour,” and “treatment gain,” along with “sex offender.” Abstracts of the references accumulated in this way were then examined, and articles that appeared to involve some assessment of change achieved during treatment were collected. Reference lists from these articles and lists of articles subsequently citing them (linked within the database) were also scanned for other relevant sources.

Table 1 outlines the 22 studies yielded from the search, ranging in date of publication from 1994 to 2007. The studies were classified according to the method of assessing treatment outcome utilised, and are listed in the table under the following

Table 1
Review of Studies Incorporating Assessment of Treatment Outcome

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
CHANGE ON DYNAMIC RISK			
Quinsey, Khanna, & Malcolm (1998)	Evaluation of a treatment programme. Also explored relationship between progress in treatment and recidivism.	Administered the following psychometric tests pre- and post-treatment: Personal Reaction Inventory – Lie; Hostility; Psychopathy; Aggression Buss-Durkee Hostility and Hostility Towards Women Marlowe-Crowne Social Behavior (social desirability in responding) Test of Attentional and Interpersonal Style – Negative Affect; Behavioral Control; Cognitive Control Adult Self-Expression Scale (assertiveness) Thorne Sex Inventory – Loss of Control Inmate Sexual Knowledge Test	Programme ineffective at reducing sexual and violent recidivism. Significant changes across treatment on majority of tests; few tests related to recidivism; post-treatment scores less so than pre-treatment. Relationship between psychometric change and recidivism not analysed.
Bakker, Hudson, Wales, & Riley (1998)	Evaluation of a treatment programme for sexual offenders against children (Kia Marama)	Psychometric battery administered pre- and post-treatment, assessing the following three domains: Sexual Attitudes and Beliefs – Abel & Becker Cognitions Scale; Hostility Towards Women; Rape Myth Acceptance Scale; Wilson Sex Fantasy Questionnaire Emotional Functioning – Beck Depression Inventory; State-Trait Anxiety Inventory; State-Trait Anger Expression Inventory Interpersonal Competency – Social Self-Esteem Scale; Assertion Inventory; UCLA Loneliness Scale; Social Avoidance and Distress Scale; Self-Efficacy Scale; Interpersonal Reactivity Index	Significant prosocial changes on majority of tests. For certain factors, non-recidivists had prosocial changes while recidivists showed the opposite , e.g.: conservative attitudes towards women, deviant fantasies, trait anxiety, anger suppression, empathy deficits.
Hudson, Wales, Bakker, & Ward (2002)	<i>Relationship between dynamic factors and recidivism (same sample as Bakker et al., 1998).</i>	<i>Same test battery as Bakker et al. (1998) administered pre- and post-treatment. Change scores calculated by subtracting post-treatment scores from pre-treatment.</i>	<i>Prosocial change generally related to reduced recidivism; correlations small and variable. Some contradictory findings.</i>

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
Marshall et al. (2002); and Marshall, Serran, Fernandez, Mann, & Thornton (2003)	Effect of therapist characteristics on indices of behaviour change across treatment.	The following treatment targets assessed pre- and post-treatment using 44 unspecified psychometric measures: Relationship quality Denial of responsibility and harm Denial of planning/premeditation Victim blaming Minimisation of features of offence Rape myth acceptance Overall treatment benefits Correlations between change scores (pre-post difference) and ratings of therapist features examined.	Therapist features (empathy, warmth, rewarding, directive) related to beneficial change in treatment. Relationship between treatment change and recidivism not reported.
Bickley & Beech (2003)	Change in distorted beliefs across treatment for “approach” and “avoidant” child sex offenders.	Measured distorted beliefs about sexual activity with children and distorted beliefs about their own victims pre- and post-treatment using the following psychometric scales: Cognitive Distortions Scale (Children & Sex Cognitions Scale) Victim Empathy Distortions Scale	Reduction in distorted beliefs across treatment for the “approach” group. Recidivism information not reported.
Marques, Wiederanders, Day, Nelson, & van Ommeren (2005)	Prospective evaluation of the effectiveness of a Relapse Prevention programme. Randomised control trial comparing recidivism of treated offenders with untreated volunteers and refusers. Also compared within treated sample based on treatment performance.	Multiphasic Sex Inventory administered pre- and post-treatment. Change on the following subscales was assessed: Cognitive Distortions and Immaturity Justifications Phallometric assessment also conducted pre- and post-treatment to assess changes in deviant sexual arousal towards scenes depicting female children, male children, and rape.	Significant pre-post treatment changes found for all scales. Re-offenders higher on deviant arousal, but relationship between change and recidivism not reported.
Beech & Hamilton- Giachritsis (2005)	Relationship between therapeutic climate and treatment outcome.	Assessed whether individual offenders’ post-treatment scores on the following three self-report measures of pro-offending attitudes were indistinguishable from non-offenders (“clinically significant change”): Victim Empathy Distortions Scale Cognitive Distortions Scale (Children & Sex Questionnaire) Emotional Identification with Children (Children & Sex Questionnaire)	Overall, 66% of the sample showed clinically significant change, variable according to group cohesiveness and expressiveness. Relationship between change and

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
Beech & Ford (2006)	Relationship between risk, deviance, treatment outcome, and sexual reconviction among child sexual abusers.	Assessed whether men in the sample had made statistically reliable changes in their attitudes following treatment to scores below the cutoff between a sample distribution of child abusers sample and a non-offending distribution (i.e., clinically significant and reliable change). Used the following psychometric measures pertaining to the two domains of deviance as defined by Beech (1998): Pro-offending Attitudes – Sexual Interest in Children (Children & Sex Questionnaire); Emotional Identification with Children (Children & Sex Questionnaire); Victim Empathy Distortions Socio-affective Functioning – Short Self-Esteem Scale; UCLA Emotional Loneliness Scale; Under-assertiveness/Over-assertiveness Scales (Social Response Inventory); Personal Distress Scale (Interpersonal Reactivity Index); Nowicki-Strickland Locus of Control Scale	recidivism not reported. None of the 16 men identified as having responded to treatment were reconvicted, compared to 5 out of 35 (14%) non-responders.
RISK ASSESSMENT TOOLS – incorporating treatment outcome			
Browne, Foreman, & Middleton (1998)	Prediction of treatment drop-out from risk factors listed in the Structured Anchored Clinical Judgement (SACJ).	Examined risk factors including those from the SACJ (Thornton, 1997, cited in Browne et al.). These related to background and criminal history, current offence, current situation and psychological problems, and the following yes/no items regarding treatment process (rated from file reports): Delinquent during treatment Deterioration in treatment No improvement in dynamic risk factors No improvement at all	Delinquent behaviour and deterioration in treatment items “medium predictors” of dropout. Dropout predictors also related to whether improvement in dynamic risk factors. Recidivism not reported.
Barbaree, Seto, Langton, & Peacock (2001)	Evaluated the predictive accuracy of 6 risk assessment instruments, including the Multifactorial Assessment of	MASORR is a guided clinical approach to risk assessment. Ratings are made on a 0-5 scale on several items empirically linked to recidivism (e.g., psychopathy, deviant sexual interests as measured phallometrically, offence history) then combined subjectively by the therapist to form a global pre-	Poor inter-rater reliability. Pre- and post-treatment scores not significantly related to sexual or serious

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
	Sex Offender Risk for Recidivism (MASORR), which incorporates treatment outcome.	treatment risk rating. After treatment, this rating is subjectively combined with ratings on the following treatment-related variables to form a global post-treatment risk rating from 1-5: Motivation for treatment Degree of behaviour change achieved Clinical impression of risk	recidivism.
Olver, Wong, Nicholaichuk, & Gordon (2007)	Development and validation of the Violence Risk Scale: Sex Offender version (VRS:SO) – a treatment and risk assessment measure comprising both static and dynamic factors and incorporating treatment change.	VRS:SO consists of a 7-item static scale, and a 17-item dynamic scale including the following 3 factors: Sexual Deviance – Sexually deviant lifestyle; Sexual compulsivity; Offence planning; Sexual offending cycle; Deviant sexual preference Criminality – Criminal personality; Interpersonal aggression; Substance abuse; Community support; Impulsivity; Compliance with community supervision Treatment Responsivity – Cognitive distortions; Insight; Treatment compliance; Released to high risk situations; and 2 non-loading items – Emotional control; Intimacy deficits Dynamic factors rated on 0-3 scale with higher scores indicating treatment needs. Motivational ratings also made for each item pre-treatment, based on the Transtheoretical Change Model (precontemplation, contemplation, preparation, action, maintenance). Motivational ratings are repeated following treatment; post-treatment scores are calculated by subtracting 0.5 for each stage progressed.	Good reliability and concurrent validity. Pre- and post-treatment scores related to sexual recidivism, and after controlling for static risk. Change scores significantly related to recidivism controlling for static risk, pre-treatment dynamic scores, and time at large. Correlations between change and recidivism significant only for high risk offenders.

SYSTEMATIC CLINICAL RATINGS – of overall treatment outcome

Hogue (1994)	Development of Standard Goal Attainment Scaling (SGAS) – a structured system for measuring the observed clinical progress of sex offenders in treatment.	SGAS consists of 12 items based on common treatment targets of cognitive-behavioural programmes, including clinical and participation dimensions: Clinical – Acceptance of guilt of the offence; Show insight into victim issues; Show empathy for their victims; Acceptance of personal responsibility; Recognising cognitive distortions; Minimisation of consequences; Understanding of life-styles dynamics; Understanding	Good inter-rater reliability, face validity and convergent validity. Recidivism information not reported.
--------------	--	---	--

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
		<p>offence cycle; Identification of relapse prevention concepts</p> <p>Participation – Disclosure of personal information; Participation in treatment; Motivation to change behaviour</p> <p>Each item rated according to behavioural descriptors on a 5-point scale from -2 to 2, with score of zero reflecting a minimally acceptable level of goal attainment. Scores can be summed to form a total. SGAS can be used to monitor progress across treatment.</p>	
<i>Stirpe, Wilson, & Long (2001)</i>	<i>Measured clinical impact post-treatment and at community follow-up.</i>	<i>Used SGAS (above) pre-treatment, post-treatment, and at 3-month follow-up in the community to track treatment gains and maintenance.</i>	<i>Clinical improvement shown across treatment. Recidivism not reported.</i>
<i>Anderson, Gibeau, & D'Amora (1995)</i>	Development and reliability data of the Sex Offender Treatment Rating Scale (SOTRS) - a treatment process and outcome measure.	<p>SOTRS is designed to be completed by therapists at various points throughout treatment to assess progress. A 6-point ratings scale is used assisted by behavioural descriptors for 54 items relating to the following dimensions:</p> <p>Insight – understanding of personal offence motives, beliefs and attitudes</p> <p>Deviant Thoughts – and offence related sexual fantasies and impulses</p> <p>Awareness of Situational Risks – which challenge capacity for self-control</p> <p>Motivation – for personal change through treatment</p> <p>Victim Empathy – emotional impact of sexual offences</p> <p>Offence Disclosure</p> <p>Ratings are combined for an overall treatment progress score, and a Progress Estimate is also made on an undefined 10-point scale from “little or no progress” (1) to “ready to leave treatment” (10).</p>	Good internal, inter-rater, and test-retest reliability reported. Recidivism information not included.
<i>Levenson & Macgowan (2004)</i>	<i>Relationship between engagement, denial, and treatment progress.</i>	<i>Used the SOTRS (see above) to assess treatment progress as the outcome measure of this study.</i>	<i>Recidivism not reported.</i>
<i>Quinsey, Khanna, & Malcolm (1998)</i>	Evaluation of a treatment programme. Also explored	<p>Therapists rated treatment gains on a 4-point scale:</p> <p>Poor – few/very few gains outweighed by deficits still requiring</p>	Programme ineffective at reducing sexual and violent

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
	relationship between progress in treatment and recidivism.	treatment, very qualified or no recommendation for gradual release, psychological treatment prior to release recommended Fair – no description given Good – no description given Very Good – significant gains in all areas targeted, no major areas requiring further work, recommendation of early release readiness, no recommendation for psychological treatment post-release	recidivism. Therapist ratings of treatment gains were not associated with reoffending.
Seto & Barbaree (1999)	Relationship between psychopathy, treatment behaviour, and recidivism.	Rated the following retrospectively from file information: Behaviour in Group - disruptiveness of conduct; appropriateness of interactions; attendance (tardiness or absence); helpfulness to others; level of participation Treatment Change - change in victim empathy; understanding of offence cycle; quality of relapse prevention plan Ratings by the therapist and clinical director on “Motivation for treatment” and “Overall degree of change” were also found in the files. All items were dichotomised (1 being positive and 0 being neutral or negative) and then averaged to form a “Treatment Behaviour” score for each offender.	Positive treatment behaviour paradoxically related to increased serious recidivism (sexual and/or violent).
<i>Barbaree (2005)</i>	<i>Addressed the same research question as Seto & Barbaree (above), using updated and more thorough recidivism data for the same sample</i>	<i>Used the same ratings of treatment behaviour as Seto & Barbaree (above)</i>	<i>In this reanalysis, there was no relationship between treatment behaviour and recidivism.</i>
Scalora & Garbin (2003)	Relationship between static risk factors, treatment-related variables, and recidivism.	Discharge reports examined for evidence of treatment involvement and level of success in attaining programme goals (responsibility for offence, victim empathy, arousal reconditioning, relapse prevention) – Sample classified as: “Successfully treated” – those who completed goals “Unsuccessfully treated” – limited progress on goals	Successfully treated offenders significantly less sexual recidivism (3%) than unsuccessfully treated (23%); risk not controlled.
Seager, Jellicoe, & Dhaliwal (2004)	Treatment efficacy study comparing recidivism of	Dichotomous pass/fail ratings for each of the following treatment components based on clinical judgement guided by a checklist:	Programme completers less likely to reoffend (sexually

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
	treatment completers with dropouts and refusers. Also compared within completers based on therapy team's evaluation of success or otherwise.	Disclosure – accepting accountability for sexual crimes Victim Empathy – emotional and cognitive recognition of harm to victims Masculinity/Femininity – understanding role of hypermasculinity and patriarchy in sexual offending Sexuality – understanding of human sexuality and development of specific sexual behaviours Relapse Prevention – understanding of concepts and formulation of plan Final evaluation also a dichotomous pass/fail judgement, based on evidence of clinical progress or remaining clinical concerns.	and/or violently) than dropouts or refusers. Within completers group, evaluations of treatment change unrelated to recidivism.
Marques, Wiederanders, Day, Nelson, & van Ommeren (2005)	Prospective evaluation of the effectiveness of a Relapse Prevention programme. Randomised control trial comparing recidivism of treated offenders with untreated volunteers and refusers. Also compared within treated sample based on treatment performance.	Participants scored on a “Got it” scale of treatment outcome consisting of the following post-treatment measures, reflective of the programme goals: Scores on four Multiphasic Sex Inventory subscales (Cognitive Distortions and Immaturity, Justifications, Child Molest, and Rape) Phallometric assessments of deviant sexual interests (male children, female children, and rape) Therapist ratings on a 7-point scale regarding the quality of two in-treatment Relapse Prevention exercises. Sample split into those who “Got It” and “Did Not Get It” based on median split on this scale.	“Got It” group had lower rates of sexual recidivism (14%) than “Did Not Get It” (27%). Difference significant among high risk offenders and child molesters. “Got It” scale a significant predictor of recidivism rate after controlling for static risk.
Looman, Abracen, Serin, & Marquis (2005)	Relationship between psychopathy, treatment change, and recidivism.	Treatment reports retrospectively coded on the following dimensions (primary contemporary treatment targets theoretically related to recidivism): Victim Awareness Quality of Offence Cycle Relapse Prevention Plans Each dimension rated on 4-point scale according to behavioural descriptions. Scores summed to produce total “treatment behaviour” score. A global “risk rating” made by the therapist at the end of treatment as to whether or not each offender's risk had been reduced, taking into account pre-treatment risk assessed by actuarial tools (VRAG; Static-99), treatment performance, and behaviour in the unit and during off-unit activities.	Low agreement between “treatment behaviour” measure and “risk rating”. Treatment behaviour unrelated to recidivism. Lower recidivism among those judged as having reduced risk (25%) than no reduction (34%); difference approached significance.

<i>Study</i>	<i>Primary research question</i>	<i>Method of assessing treatment outcome</i>	<i>Results</i>
Langton, Barbaree, Harkins, & Peacock (2006)	Relationship between treatment response and recidivism as a function of psychopathy.	Revised the Treatment Behaviour scale used by Seto & Barbaree (1999). Ratings on 8 items were made on a 2 or 4-point scale according to criterion-based exemplars, creating the following subscales: Conduct – Attendance; Appropriateness; Participation (homework); Participation (group work) Treatment Targets – Denial/minimisation; Victim empathy; Offence cycle; Relapse prevention plan Scores for overall Response to Treatment were created by averaging ratings.	Response to treatment scores not significantly related to serious or sexual recidivism.

headings:

- Change on dynamic risk
- Risk assessment tools incorporating treatment outcome
- Systematic clinical ratings of overall treatment outcome

The studies are listed in Table 1 in chronological order under these headings, except where studies have utilised the same measure of treatment outcome described in a previous study, in which case they are listed directly below the first regardless of publication date (e.g., Stirpe et al., 2001, follows Hogue, 1994), and are italicised. The columns in Table 1 present the authors and publication date, the primary research question of the study, a detailed description of the method or methods of assessing treatment outcome employed in the study, and finally the pertinent results relating to treatment outcome. Results pertaining to the relationship between proximal treatment outcome and recidivism at follow-up, if reported, are shown in bold. Two studies (Quinsey et al., 1998, and Marques et al., 2005) are entered twice in the table, as they included two separate methods of assessing treatment outcome which fell under different headings.

The following section will outline the studies presented in Table 1, the overall patterns that emerged regarding the different conceptualisations and operationalisations of treatment outcome in the literature, the relationships between these and recidivism, followed by a discussion of the advantages and disadvantages of each.

Change on dynamic risk.

Several studies have assessed in-treatment change by measuring dynamic risk factors of an offender prior to treatment, and then reassessing the same factors

following completion of the programme. Typically, these pre- and post-treatment assessments involved psychometric tests designed to measure variables that had previously been empirically or theoretically linked to risk, were assumed to be dynamic (changeable), and were targeted for change in the treatment programme. Examples include cognitive distortions and pro-offending attitudes (e.g., Children and Sex Questionnaire; Facets of Sexual Offender Denial; Rape Myth Acceptance Scale), interpersonal competencies (e.g., Inmate Sexual Knowledge Test; UCLA Loneliness Scale; Adult Self-Expression Scale), victim empathy (e.g., Victim Empathy Distortions Scale), and emotional functioning (e.g., State-Trait Anger Expression Inventory). In addition to psychometric tests, one study (Marques et al., 2005) conducted pre- and post-treatment phallometric testing to assess changes in deviant sexual arousal.

Table 1 shows that it is typical for studies measuring treatment outcome in this way to find that treatment appeared to result in significantly reduced dynamic risk (i.e., change in the direction intended according to treatment goals). This averaged effect was reported by Quinsey et al (1998), Bakker et al. (1998), the “approach” group in Bickley and Beech (2003), and Marques et al. (2005). In all other studies that measured change on dynamic risk across treatment psychometrically but did not report averaged treatment change results, positive change was alluded to for at least some of the sample (Hudson et al., 2002; Marshall et al., 2002; Marshall et al., 2003; Beech & Hamilton-Giachritsis, 2005). This is encouraging, however it has been suggested that self-report psychometric tests are problematic in that they can be transparent, and therefore are open to a social desirability bias in responding (Tierney & McCabe, 2001). Pre- and post-treatment phallometric assessments (included in Marques et al., 2005, as part of the treatment change assessment) should theoretically

be more objective, since they rely on physiological responses rather than self-report. However, they are only applicable as a measure of one dynamic risk dimension (deviant sexual interests), therefore are not sensitive to overall change as a result of treatment. Additionally, non-responsiveness is common, and individuals with low responses have been shown to score higher on measures of social desirability (Looman, Abracen, Maillet, & DiFazio, 1998), indicating that phallometric assessments are not without limitations as a method of evaluating treatment change.

Therefore, although assessments of dynamic change can appear very positive, it is important to establish a link between in-treatment change on dynamic risk factors and decreased sexual recidivism. The Kia Marama reports (Bakker et al., 1998; Hudson et al., 2002), and Beech and Ford (2006), are the only to have reported the relationship between psychometric change and recidivism. Bakker et al. (1998) compared mean pre- and post-treatment scores of recidivists ($n = 19$) with non-recidivists ($n = 219$), and computed an effect size comparing the two groups' differing response to treatment. Significant differences were found on some measures (including Impersonal and Sado/Masochistic subscales of the Wilson Sex Fantasy Questionnaire [WSFQ], Trait Anger and Anger Suppression as measured by the State-Trait Anger Expression Inventory [STAXI], and the Perspective Taking and Identification with Others subscales and total scores of the Interpersonal Reactivity Index [IRI], a measure of empathic responding), such that the group who eventually reoffended sexually had on average worse scores post-treatment than pre-treatment, while the opposite was true for the non-recidivist group. For the majority of scales in the battery, however, this effect was not significant, and there were no apparent differences in the treatment response of recidivists and non-recidivists as measured in

this way. A limitation is that the size of the recidivist group may have been too small to detect effects using group comparison methodology.

Hudson et al. (2002) used a different method to analyse the relationship between treatment change and reoffending at Kia Marama. They calculated change scores (pre-treatment minus post-treatment) for each individual on each psychometric measure, and then examined the correlations between these and sexual recidivism. Across the battery, these correlations tended to be in the expected direction (prosocial change related to reduced recidivism), although relatively few reached significance. The most promising findings were in the interpersonal competency domain of the battery, with prosocial changes significantly related to reduced recidivism for assertiveness and IRI (empathy) scores. From the emotional functioning domain, lower trait anger was also significantly correlated with reduced recidivism. However, contrary to expectations, prosocial change on some subscales (Impersonal and Sado/masochistic subscales of the WSFQ, and the Suppression subscale of the STAXI) was significantly correlated with increased recidivism. A limitation of the Hudson et al. (2002) study that may be related to the disparate results on these scales is the use of raw difference scores. As noted by the authors, change scores do not account for pre-treatment level, and those who are most deviant at pre-treatment have the greatest capacity to evidence change. Therefore, correlations between prosocial change and increased recidivism may simply reflect the higher underlying dynamic risk of those whose scores indicated greater treatment change on these self-report measures.

Beech and Hamilton-Giachritsis (2005) described and utilised an alternative method of analysing psychometric assessments of progress in treatment, which involved determining whether an individual's post-treatment scores on tests

measuring dynamic risk were indistinguishable from normative scores of non-offenders. If so, the offender was deemed to have achieved “clinically significant change.” This method may potentially provide a standardised way of incorporating post-treatment psychometric data into individualised risk judgements that are easily interpretable, without being affected by the problem of variable pre-treatment scores. Standardising also allows for the reduction of data; for example, to obtain a single treatment change score rather than several scores for multiple tests. However, since clinically significant change was the outcome measure for Beech and Hamilton-Giachritsis’ study, (the specific research question being the relationship between therapeutic climate and treatment outcome), no information on recidivism was reported. Additionally, although potentially an important advancement for the practice of post-treatment risk assessments at the individual level, clinically significant change methodology is not strictly an assessment of change in treatment, as it only involves consideration of post-treatment scores against external criteria (normative means). It therefore cannot address the question of whether the dynamic risk factors that are being assessed are really dynamic, or whether specific changes in treatment are successful in reducing recidivism.

Beech and Ford (2006) applied the same methodology in an examination of the relationship between static and dynamic risk, treatment outcome, and recidivism, however they took into account the reliability of the change achieved as well as the clinical significance. Change was deemed reliable if the effect size was greater than that which would be expected due to chance. After a follow-up period of two years, none of the group identified as having responded to treatment were reconvicted, compared to 14% of those who did not respond. These results indicate further support for prosocial change on dynamic risk factors being associated with decreased

recidivism, however these findings were based on a small sample ($N = 51$) with only five identified recidivists (also, 20% of the original sample could not be identified at follow-up). It additionally seems likely that clinically significant change methodology may suffer from the same problem of socially desirable responding as do other methods of analysing self-report psychometrics.

Overall, it can be seen that as yet there is a lack of reliable and consistent findings linking in-treatment dynamic change (measured psychometrically) with decreases in recidivism, although some promising results have been reported (Beech & Ford, 2006; Hudson et al., 2002). Important directions for future research could include taking into account the level of pre-treatment deviance when examining the relationship between psychometric change and recidivism; or to apply clinically significant change methodology to dynamic risk frameworks such as that devised by Allan et al. (2007), which provide another way of standardising psychometric data to enable averaging across tests.

Risk assessment tools incorporating treatment outcome.

This section of Table 1 lists three studies involving the use of measures designed to assess risk in sex offenders that incorporate judgements of change or progress achieved in treatment. Two of these measures (Structured Anchored Clinical Judgement, SACJ; Browne, Foreman, & Middleton, 1998; and Multifactorial Assessment of Sex Offender Risk for Recidivism, MASORR; Barbaree, Seto, Langton, & Peacock, 2001) include specific items pertaining to performance in treatment or level of change achieved that must be rated by therapists completing the measure along with other items related to risk (e.g., the SACJ includes the items “deterioration in treatment” and “no improvements in dynamic risk factors,” rated on

a yes/no basis). The third measure (Violence Risk Scale: Sexual Offender Version, VRS:SO, Olver et al., 2007) incorporates a structured assessment of treatment change to adjust pre-treatment scores on 17 dynamic items according to evidence of progression through the stages of change described by the Transtheoretical Change Model (Prochaska et al., 1992). Progression through the stages for each item is judged by factors such as recognition of the problem, desire to change, relevant behavioural improvements, and consistency, stability, and generalisation of the improved behaviours (Wong, Olver, Nicholaichuk, & Gordon, 2006).

The SACJ and the MASORR are both relatively unstructured risk assessment tools that could be described as being “guided clinical” approaches to risk assessment (Barbaree et al., 2001). Browne et al.’s (1998) study simply involved making yes/no ratings on SACJ risk items and some additional items. Four treatment process variables (see Table 1) were included. Frequencies were then obtained for each item, and correlated with the yes/no treatment dropout variable (the outcome measure of the study). They reported some interesting results regarding predictors of treatment dropout (which included treatment process variables), and reported that predictors of dropout were also related to whether an offender showed an improvement across treatment in dynamic factors (yes/no judgement). However, information regarding the relationship between the treatment process variables and recidivism was not included in the study.

The MASORR, included in Barbaree et al.’s (2001) comparative evaluation of six risk assessment instruments for adult sex offenders, involves subjectively combining pre-treatment ratings of empirically validated risk factors (e.g., psychopathy, social competence, deviant sexual interests, offence history), and then following treatment subjectively combining the risk estimate with ratings of treatment

outcome factors (motivation for treatment and degree of behaviour change achieved) and a clinical impression of risk. Barbaree et al.'s (2001) evaluation study found that the MASORR had poor inter-rater reliability, neither pre-treatment nor post-treatment scores were significantly related to sexual recidivism, and the inclusion of the treatment-related items in the post-treatment assessments did not improve the predictive value of the measure. Of the individual items, the only one that reached significance was Offence History; the treatment outcome-related variables had insignificant correlation magnitudes with sexual recidivism in the range of $r = .08$ to $.15$. These results do not suggest much promise for structured clinical judgement risk tools incorporating treatment progress items.

In contrast to these two examples of guided clinical judgement risk instruments, the VRS:SO is described as merging actuarial and clinical traditions (Olver, 2003), because although dynamic item ratings are necessarily less objective than static items as they involve some clinical judgement, the VRS:SO includes structured scoring criteria to increase objectivity, and scores can be translated into probabilistic risk estimates (the VRS:SO also includes a static actuarial subscale). As discussed in the previous section of this review, the validity of the VRS:SO as a risk assessment instrument seems promising, as Olver et al.'s (2007) validation study reported good inter-rater reliability and predictive validity of the measure. They also found that the dynamic scale scores were predictive of sexual recidivism after controlling for the VRS:SO static scale and the Static-99.

Olver et al. (2007) additionally examined the relationship between treatment change as measured by the VRS:SO and sexual recidivism. Change scores were calculated by subtracting post-treatment dynamic scores from pre-treatment dynamic scores; this is also equivalent to the total number of points deducted for progression

through the stages of change across all 17 dynamic items – a reduction of 0.5 points for each stage progressed per item (according to VRS:SO coding protocols; Wong et al., 2006). They found support for the contention that changes on dynamic risk factors as a result of gains made in treatment could be associated with reductions in recidivism, at least for high risk offenders. For their entire sample, change scores were significantly predictive of recidivism after controlling for static risk, pre-treatment dynamic scores, and time at large using Cox regression. These findings are very encouraging, however the authors noted the necessity of cross-validation studies on different samples of treated sex offenders, in order to evaluate the generalisability of the results.

Olver et al. (2007) additionally found that when analysed separately, treatment change was unrelated to sexual recidivism for low risk men (defined by Static-99 scores of 0-3). They determined that changes made in treatment may be more informative for higher risk offenders, possibly a result of them having more potential to demonstrate changes due to their higher pre-treatment scores. Because the VRS:SO is primarily a risk instrument, post-treatment scores must balance the potential long-term nature of risk against the unknown stability of treatment gains. The criteria for scoring treatment change on the VRS:SO can therefore be considered relatively conservative. For example, an individual who received the maximum score of 3 for a particular dynamic item and was rated as being in the Precontemplation or Contemplation stage of change at pre-treatment could only reduce their score as low as 1.5 at post-treatment, even if they progressed to the final stage of change, Maintenance. Furthermore, in a prison-based programme progressing to Maintenance may be difficult, as the opportunity to show evidence of the stability of the positive behaviour changes across a variety of relevant high risk situations (required according

to rating protocols; Wong et al., 2006) might not arise prior to release. A post-treatment score of 2 for a particular risk factor would therefore not be uncommon for an individual who had performed exceedingly positively in treatment and made substantial gains. It may be that assessments of treatment outcome that are independent from considerations of risk, such as those reviewed in the next section, will be able to offer an additional perspective.

In summary, evidence for the validity of sex offender risk assessment instruments that include guided clinical judgements of treatment outcome (e.g., SACJ and MASORR) is poor at this stage. A risk instrument incorporating a structured theoretical approach to assessing change in treatment (VRS:SO) shows good predictive validity, and support for the association between in-treatment changes on dynamic factors and reductions in recidivism, although independent replication studies are required.

Systematic clinical ratings of overall treatment outcome.

The final section of Table 1 lists studies in which treatment outcome was assessed by therapists or researchers according to structured rating systems, usually based on behavioural descriptors. This section includes the two studies mentioned earlier in which the aim was to develop a structured, reliable, and valid method of assessing short-term treatment outcome among sex offenders – the SGAS (Hogue, 1994) and the SOTRS (Anderson et al., 1995). Both of these measures were reported to have good inter-rater reliability and were able to track the progress of sex offenders across treatment. SGAS scores were also shown to be correlated with several psychometric measures relating to sexual aggression (including the Attitudes Towards Sex With Children scale, and the Justifications and Cognitive Distortions and

Immaturity subscales of the Multiphasic Sex Inventory). The SGAS and the SOTRS therefore show promise as measures of treatment progress, but neither has yet been validated as being predictive of sexual recidivism – this could be an important advancement for future research.

Aside from these two measures, eight other studies have described idiosyncratic systems of rating short-term treatment outcome that fit within this category of Table 1. These systems varied considerably by level of complexity. For example, Scalora and Garbin (2003) and Looman, Abracen, Serin, and Marquis (2005) employed relatively simplistic dichotomous ratings (e.g., “successfully treated” or “unsuccessfully treated”), while other studies rated subjects on up to eight dimensions (Seto & Barbaree, 1999; Langton, Barbaree, Harkins, & Peacock, 2006). The rating systems also vary by content. Some studies rated clinical change only (Seager et al., 2004), while others also included items pertaining to behavioural conduct in treatment (Seto & Barbaree, 1999; Langton et al., 2006). Marques et al.’s (2005) “got it” scale incorporated post-treatment psychometric and phallometric scores as well as therapist ratings of the quality of treatment exercises. The inclusion of this study under the *Systematic Ratings* heading of Table 1 is therefore debatable, but it was put there since change on these measures was not assessed (and so it did not belong under the *Change on Dynamic Risk* heading), and the measures were selected as being relevant to the treatment goals of the programme.

All eight of the studies reported the relationship between their measures and recidivism. Due to the paucity of these types of analyses for other methods of assessing treatment change, the majority of reported findings on the relationship between treatment gains and recidivism therefore fall within this category (systematic clinical ratings).

Several studies reported that no relationship was found between their assessments of treatment outcome and recidivism at follow-up. In a study by Quinsey et al. (1998), therapists rated treatment gains on a simple 4-point scale from “poor” to “very good” for 193 men who completed a sex offender treatment programme as part of an evaluation of the efficacy of the programme. Comparisons of the recidivism rates of their sample who received treatment with those of an untreated control sample led Quinsey et al. (1998) to conclude that their programme was ineffective in reducing sexual or violent recidivism. Consistent with this, scores on the scale of treatment gains among those who received treatment were unrelated to recidivism. However, an important possible confound is that criteria for a rating of very good treatment gains included if a recommendation had been made for early release readiness. This suggests that those rated as having performed better in treatment may have been released earlier, yet time at large does not appear to have been controlled for in the recidivism analyses.

An influential study by Seto and Barbaree (1999) reported the results of an examination into the relationship between psychopathy, treatment behaviour, and recidivism among 216 sex offenders. As discussed in the previous section of this review, psychopathy has been identified as an important risk factor for sex offenders, consisting of traits such as callous disregard for the rights of others, manipulateness, and impulsivity (Hare, 1991). Psychopathy has also been identified as a responsibility issue, meaning that its presence could impact on an offender’s ability to respond to treatment beneficially (e.g., Rice et al., 1992). Seto and Barbaree (1999) were therefore interested in whether an interaction effect would be found, such that the presence of psychopathic traits would influence any relationship between treatment behaviour and recidivism. Their measure of “treatment behaviour” included a

combination of treatment performance related items such as disruptiveness and participation, and items pertaining to clinical outcome such as change in empathy and understanding of offence chain, and therefore is relevant to this review of treatment outcome measures. Unlike Quinsey et al. (1998), Seto and Barbaree (1999) did find a significant relationship between treatment outcome and recidivism, however this relationship was in the opposite direction to that expected. Positive evaluations of treatment outcome were related to increased, rather than decreased, serious recidivism (a combined outcome of sexual and/or violent reoffending). They additionally reported an interaction effect, whereby the combination of high psychopathy and good treatment behaviour (defined as having scores above the sample mean for both measures) was associated with a serious recidivism rate more than five times that of the remainder of the sample.

As noted, Seto & Barbaree's (1999) results were influential, the implications pertaining not only to the appropriateness of providing this kind of treatment to sex offenders with high levels of psychopathic traits, but also raising the worrying possibility that appearing to do well in a treatment programme in terms of motivation, participation, and positive change, might actually be a factor indicative of increased risk. However, six years later, Barbaree (2005) published a re-analysis of the same data, but applying a longer follow-up period and more complete recidivism data from a national police database. In contrast to the earlier findings, Barbaree's (2005) re-analysis indicated that there was no statistically significant interaction effect between psychopathy and treatment behaviour on recidivism, and indeed, the measure of treatment behaviour was unrelated to recidivism. This was further supported in a study by Langton et al. (2006), who applied a modified version of Seto and Barbaree's (1999) treatment behaviour measure to obtain an averaged "Response to

Treatment” score, which was found to be unrelated to serious or sexual recidivism. Finally, Seager et al. (2004) also found no relationship between recidivism and treatment outcome (therapist judgements of “pass” or “fail” for five treatment targets and a final evaluation) among 76 men who completed a prison-based sex offender treatment programme.

Looman et al. (2005) conducted a study in which two separate measures of treatment outcome were employed on a sample of 154 male sex offenders admitted for treatment. The first was termed “treatment behaviour,” although it consisted only of items pertaining to clinical gains achieved on treatment goals (victim awareness, quality of offence cycle, and relapse prevention plans). Treatment behaviour was found to have no relationship with recidivism, supporting the majority of the studies discussed so far under this heading of Table 1. By contrast, a separate measure of treatment outcome employed by Looman et al. (2005) – a dichotomous “risk reduction” rating – was related to recidivism in the direction that would be expected, with the difference between the two groups approaching significance. Accordingly, those offenders who were judged to have reduced their risk across treatment, with reference to pre-treatment actuarial risk, treatment performance, and observed behaviour, were less likely to have a serious reconviction (combined sexual and/or violent recidivism) than those judged as having no reduction (rates of 25.3% and 34.1% respectively).

This finding has been supported by the results of two other studies, in which positive ratings of treatment outcome obtained using relatively simple scales were related to reduced recidivism. Scalora and Garbin (2003) found that out of 76 child molesters who received intensive inpatient group and individual cognitive-behavioural treatment, those who were deemed “successfully treated” in terms of

attaining treatment goals were less likely to reoffend sexually than those who made only limited progress on goals (“unsuccessfully treated”). The recidivism rates were significantly different, 3% and 23% respectively, however it should be noted that the recidivists in their study had indications of higher risk relative to those who did not reoffend. Finally, Marques et al. (2005) constructed a “got it” scale on which their sample of sex offenders treated in a relapse prevention programme as part of the Californian randomised control trial were rated. This scale was based on various factors including post-treatment psychometrics scores, phallometric assessments, and therapist ratings of the quality of treatment work, and was then reduced to a dichotomy (“got it” or “did not get it”) based on a median split. As discussed, results of the RCT were not supportive of this programme being effective at reducing recidivism. Nonetheless, more promising results were found in the within-treatment got it analyses, as the got it group had a decreased rate of sexual recidivism (14%) relative to those who did not get it (27%), a difference that was statistically significant among those who were at high risk of reoffending according to static measures. Additionally, the got it scale was found to predict recidivism once static risk was controlled for. These results indicate some support for the efficacy of the treatment programme, and suggest the failure of the RCT to find a treatment effect may have been contributed to by methodological factors rather than solely indicating the ineffectiveness of the programme (methodological issues that may have contributed to the RCT failure were discussed at length by Marques et al., 2005).

Overall, it can therefore be seen that results regarding the relationship between treatment outcome as measured using structured clinical rating systems and recidivism are variable. However, with the exception of the findings reported by Seto and Barbaree (1999) which have since been effectively retracted following Barbaree’s

(2005) reanalysis, all of the studies report either no relationship or an association between positive treatment outcomes and reduced recidivism.

Several different conclusions could be drawn from findings indicating a lack of relationship between a measure of within-treatment outcome and recidivism. The most obvious is simply that proximal treatment gains are unrelated to reduced recidivism in the long-term, suggesting possibly that the wrong targets are being focused on, or that so called dynamic factors are not actually amenable to change in a way that is linked to recidivism. However, if a measure lacks predictive validity with regards to recidivism it is difficult to advocate its use as a measure of in-treatment progress, and so another possible conclusion to draw from such results is that the measure of treatment outcome applied is invalid. This possibility is supported by the fact that varied results have been found even within the same study using multiple operationalisations of treatment outcome (e.g., Looman et al., 2005; Marques et al., 2005), indicating that the method used to assess treatment outcome can have a significant effect on results. It therefore appears very important to ensure the validity of measures of treatment outcome in terms of their relationship with recidivism before they are applied clinically or to address other research questions.

Summary

The above review has outlined the different ways that the concept of proximal treatment outcome has previously been conceptualised and measured in the literature, falling under the three main categories of change on dynamic risk, risk assessment tools incorporating treatment outcome, and systematic clinical ratings of overall treatment outcome.

Using psychometric tests to measure across-treatment changes on dynamic risk variables has the advantages of being able to select measures with pre-established reliability and validity that are pertinent to the specific domains targeted in a particular treatment programme. Such tests also do not rely on the judgement of therapists, which may be biased by irrelevant factors (e.g., how likeable an offender being assessed is, or cases in which an offender is “faking good,” or behaving in ways that inaccurately lead a therapist to believe they are performing well in treatment). However, self-report psychometrics (and even possibly phallometric assessments) may be subject to a self-presentation bias. Using raw change scores is also problematic due to pre-treatment variance. Controlling pre-treatment scores or applying reliable and clinically significant change methodology when analysing psychometric change may be the most judicious approaches to analysing treatment outcome using change on dynamic risk.

In terms of risk assessment tools that incorporate a measure of change achieved in treatment, the VRS:SO is the only currently available tool that shows promise in terms of being related to recidivism. Further independent validation studies to test the generalisability of the promising results reported by the VRS:SO authors (Olver et al., 2007) would at this stage be an important contribution to the area of assessing treatment outcome among sex offenders.

Most of the research findings to date linking proximal treatment changes with recidivism have come from studies falling under the third heading of Table 1; that is, using post-treatment clinical ratings to assess treatment change. However, it has been shown that results are variable, and appear to be very dependant on the quality of the measure of treatment outcome employed. The measures previously applied also tend to be idiosyncratic, developed for the purpose of addressing a particular research

question and applied in a single study only. By contrast, SGAS was developed specifically for the purpose of measuring gains made by individual sex offenders in treatment programmes, and has been previously found to have good inter-rater reliability, face validity, and convergent validity, as well as the ability to track clinical improvement across treatment. SGAS would also seem to carry an advantage of efficiency over the 54-item SOTRS developed for the same purpose. Assessing the predictive validity of the SGAS in terms of predictive validity could also be an important contribution to this area.

In summary, there are numerous clinical and research applications of measuring proximal treatment outcome among sex offenders, and it is important that these estimates are linked with the longer-term outcome of recidivism. There are three general approaches to assessing treatment outcome, however research to date exploring the relationship between these and recidivism is limited with variable results. Further research is needed to assess the comparative validity of existing approaches, develop ways to improve the measurement of treatment outcome, and explore the relationship between proximal treatment gains and reduced recidivism.

Outline of the Empirical Section

The preceding literature review has given a broad overview of the historical development, current practice, and effectiveness of the treatment of sex offenders; outlined the practice of risk assessment for this population alongside the inherent challenges of such a task; explored in depth the attempts that have been made to date by researchers and clinicians to conceptualise and operationalise the concept of treatment outcome – the amount of gain or progress made during treatment; and finally, outlined previous results regarding the relationship between treatment outcome and recidivism. Several disparities in the literature and other important areas that could benefit from further research have been highlighted, and the empirical section of this dissertation to follow aims to address many of these. This dissertation contains four separate studies; the rationale for each based on the literature review and outline of the research question and methodology are introduced below.

Study 1 is an independent validation study of the Violence Risk Scale: Sexual Offender Version (VRS:SO; Olver, Wong, Nicholaichuk, & Gordon, 2007), a dynamic actuarial measure for assessing risk among sexual offenders. As noted, this measure is reflective of the latest developments in risk assessment, as it is based on a theoretical and empirical framework, and includes protocols for assessing changes made in treatment. The rationale for Study 1 is that although the predictive validity and other properties of the VRS:SO were supported by Olver et al. (2007), as noted by the authors further research is required to assess the replicability of these findings, and the generalisability of the measure to different samples and jurisdictions. In Study 1 the VRS:SO is rated retrospectively from file information of 218 child molesters who received treatment at the Kia Marama Special Treatment Unit at Rolleston

Prison, near Christchurch, New Zealand, between 1993 and 2000. The retrospective nature of this design allows for efficient access to recidivism data over adequate follow-up periods after release from prison, while the integrity of the data is preserved by the researchers being blind to recidivism outcome during coding. This design has been termed “retrospective-prospective.” This design is employed in all four studies of this dissertation, and the same core sample of 223 Kia Marama completers is used, with minor variations depending on data availability.

In the preceding section of the literature review, various methods of assessing proximal outcomes of sex offender treatment (i.e., the gains or degree of change achieved by individual offenders through completing a treatment programme) that have been employed in previous research were outlined. These methods were classified into three categories: change on dynamic risk (usually measured psychometrically); risk assessment tools incorporating treatment outcome (with the VRS:SO showing the most promise of the three reviewed); and systematic clinical ratings of overall treatment outcome. Study 2 is a comparative validity study of three such methods, one from each of the identified categories. The first involves analyses of change on a battery of psychometric tests encompassing a dynamic risk framework devised by Allan, Grace, Rutherford, and Hudson (2007). This battery is routinely administered to Kia Marama participants both before and after treatment. The second method is change scores on the VRS:SO risk instrument, previously supported by Olver et al. (2007) as being associated with reduced sexual recidivism. And thirdly we also retrospectively rated the same archival files (while blind to recidivism outcome) on a modified version of the Standard Goal Attainment Scaling measure (SGAS; Hogue, 1994), designed to assess the extent to which an offender has attained specific

goals of treatment. The relationship between each of these measures and sexual recidivism was also assessed.

The rationale for Study 2 is based on the need for validated measures of treatment outcome that are empirically related to recidivism, given the numerous clinical and research applications identified in the literature review. As outlined, despite the identified need (relevant to improving risk assessments, treatment efficacy research, responsiveness research, and so on), only a small number of published studies have incorporated measurements of treatment outcome among sex offenders (these studies were overviewed in Table 1), and even fewer have analysed the relationship between these measures and recidivism. Those measures for which the association with recidivism was assessed mostly involved idiosyncratic rating systems designed for the purposes of the specific research question and not pre-validated. Interpreting results such as a lack of relationship between treatment outcome measures and recidivism (e.g., as reported by Barbaree, 2005; and Quinsey, Khanna, & Malcolm, 1998) can therefore be problematic, as such findings could conceivably be due to the use of an inappropriate or invalid measure of treatment outcome. The aim of Study 2 is thus to evaluate the comparative validity and utility of the above three methods of assessing proximal treatment outcome, each selected as showing potential value based on previous research and being representative of the three categories of Table 1. It is hoped that this investigation will provide some clarity as to valid measures of treatment outcome, the relationship between proximal treatment outcome and recidivism, and relative advantages and disadvantages of each approach.

Studies 3 and 4 investigate the influence of particular offender characteristics on treatment outcome and reoffending, as well as the potential for factors to interact with each other in the prediction of recidivism. As discussed, variable recidivism rates

among sex offenders who have completed a treatment programme may be an effect of evaluating treated offenders as a single group (Bickley & Beech, 2003). In reality, it seems likely that a complex interplay of numerous factors internal and external to the offender would impact on both treatment response and the likelihood of reoffending. Exploring such factors could help to more effectively target the differential needs of individual offenders (i.e., address the issue posed by Marques, 1999, of “what works for what types of offenders?”), as well as refine risk assessments among this population. There are several possibilities worthy of exploration within this issue, but one that appears particularly pertinent is the personality construct of psychopathy. Psychopathy has been robustly supported as a risk factor for sexual reoffending (Doren, 2004; Hemphill, Hare, & Wong, 1998), and is also thought to be a responsivity factor (i.e., a characteristic that influences response to treatment; e.g., Rice, Harris, & Cormier, 1992). Another potentially relevant variable is intelligence. Intelligence has long thought to be associated with criminality (e.g., Hirschi & Hindelang, 1977), and has additionally been suggested to influence the relationship between psychopathy and offending (e.g., Heilbrun, 1979, 1982; Walsh, Swogger & Kosson, 2004).

Study 3 is an attempted replication and extension of Seto and Barbaree’s (1999) reported interaction effect between psychopathy and treatment outcome on recidivism. SGAS is used as the measure of treatment outcome, and psychopathy is assessed using the Psychopathy Checklist-Revised (Hare, 1991). Study 4 investigates the possible interaction between psychopathy and intelligence on sexual recidivism, further to a series of early investigations by Heilbrun (1979, 1982) regarding violence. For both studies, the potential interactive effects are explored using regression analyses, as well as Kaplan-Meier survival analyses to compare the recidivism rates

of four groups formed by performing median splits of the variables in question. Also in both studies, to further explore potential sub-sample differences, the hypotheses are explored separately for the incest and extrafamilial offenders in the sample.

Following the four-study empirical section of this dissertation there will be a general discussion, in which the results will be summarised. Overall conclusions will be outlined within the following general themes: assessment of treatment outcome among sex offenders; incorporating treatment outcome into sex offender risk assessments; and the influence of specific offender factors on assessments of treatment outcome and risk.

Study 1. Assessment of Dynamic Risk Factors: An Independent Validation Study of
the Violence Risk Scale: Sexual Offender Version

1.1 Abstract

The Violence Risk Scale: Sexual Offender Version (VRS:SO; Olver, Wong, Nicholaichuk, & Gordon, 2007) is a rating scale designed to assess risk among sexual offenders and the degree of change achieved in treatment. The scale consists of 7 static and 17 dynamic risk items, and protocols for measuring treatment change draw on the Transtheoretical Change Model (Prochaska, DiClemente, & Norcross, 1992). The aim of the current study was to evaluate the psychometric properties of the VRS:SO on an independent sample of 218 child molesters who received treatment at a prison-based program in New Zealand. Supporting the initial validation of the VRS:SO by Olver et al. (2007), our results indicated good inter-rater reliability, concurrent validity, and predictive validity of the measure – VRS:SO scores were predictive of sexual recidivism and the dynamic scale made significant incremental contributions controlling for static risk. We also analysed the validity of the factor structure of the VRS:SO dynamic scale, and compared the measure with an alternative measure of dynamic risk (Allan, Grace, Rutherford, & Hudson, 2007) based on a self-report psychometric battery.

Sex offender risk assessments inform many decisions across the criminal justice process, such as sentencing, security ratings, treatment indications, whether and when to release, and probationary conditions. These decisions have important real-life implications for many parties which need to be balanced against each other, including the offender and their family, the wider community, and potential future victims (Doren, 2006). The validity of risk assessments is therefore of the utmost importance, and ongoing research is essential to improve their accuracy.

Modern risk assessments are actuarial in nature, based on objective evaluation of factors empirically related to recidivism. The development of actuarial risk assessments was motivated by evidence that unguided clinical judgments generally offered poor predictive validity compared to more structured methods (e.g., Grove & Meehl, 1996). Variables which are predictive of recidivism are known as risk factors, and can be classified as either static or dynamic (Andrews & Bonta, 2003). Static factors are unchangeable, often historical or offense-related, such as number of previous sexual offenses. Dynamic factors are those variables which are related to recidivism but are changeable, at least in principle, and thus can potentially serve as targets for intervention. Examples of dynamic risk factors for sexual offending include sexual deviancy, antisocial orientation, poor social functioning, and distorted cognitions (see Craissati & Beech, 2003, for review).

Meta-analyses have provided substantial evidence that both static and dynamic factors are valid risk predictors for sexual reoffending. Hanson and Bussière (1998) identified several primarily static risk factors with significant and reliable correlations with sexual recidivism, including: young age; single marital status; antisocial personality; prior sexual and non-sexual offenses; extrafamilial, stranger, or male

victims; deviant sexual interests; and treatment non-completion. Subsequently, Hanson and Morton-Bourgon (2005) identified additional dynamic risk factors that have potential as treatment targets, including sexual deviance, antisocial orientation, self-regulation problems, and distorted cognitions.

Over the last decade there has been much progress in the development of actuarial risk prediction instruments. These instruments have been based most heavily on static risk factors, because of their superior predictive validity and relatively objective nature compared to dynamic factors – for example the Static-99 (Hanson & Thornton, 1999) and the SORAG (Quinsey, Harris, Rice, & Cormier, 1998). Actuarial static risk measures have been shown consistently to predict sexual recidivism, with areas under the receiver operating characteristic curve (AUCs) for the Static-99 in the range of .61 to .76 (Barbaree, Seto, Langton, & Peacock, 2001). However, a major disadvantage of static risk instruments is that being based on historical factors, they are not sensitive to change (e.g., reductions in risk as a result of treatment). Because they do not include dynamic factors, these instruments are also not able to inform as to appropriate treatment targets, or criminogenic needs (Andrews & Bonta, 2003). Identifying treatment targets and evaluating treatment change are both important goals of real-world risk assessments beyond mere prediction of the likelihood of recidivism.

The Violence Risk Scale: Sexual Offender Version (VRS:SO; Olver, 2003) was designed as a comprehensive actuarial risk instrument for sexual offenders, which includes both static and dynamic risk factors that are empirically or theoretically related to sexual recidivism. It was derived from the Violence Risk Scale (VRS; Wong & Gordon, 2006), which had been previously developed to assess the risk of violent offending. The VRS:SO has strong theoretical foundations, incorporating both the risk-needs-responsivity approach to offender rehabilitation (Andrews & Bonta,

2003), and elements of the Transtheoretical Change Model (TCM; Prochaska, DiClemente, & Norcross, 1992). As well as predicting recidivism risk, the VRS:SO was also designed to integrate risk assessment with treatment by identifying criminogenic needs as potential treatment targets, to evaluate readiness to change on each identified need according to the TCM (an important responsivity consideration), and to assess changes in risk following treatment. The VRS:SO has the potential to make a significant contribution to the practice of risk assessment of sexual offenders, by drawing together and applying relevant theoretical bases, and providing the means to assess dynamic factors and treatment change alongside static risk.

The developers of the VRS:SO carried out initial analyses of the psychometric properties of the measure on a sample of 321 males who completed a treatment program for sexual offenders in a maximum-security forensic unit (Olver, Wong, Nicholaichuk, & Gordon, 2007). The inter-rater reliability, concurrent validity, and predictive validity of the measure were evaluated, as well as the relationship between change on the dynamic items and sexual recidivism. Additionally, an exploratory factor analysis was conducted on the 17 dynamic items, from which an orthogonal 3-factor structure was suggested: Sexual Deviancy, Criminality, and Treatment Responsivity. Inter-rater reliability was good, with an intraclass correlation coefficient (r_{ICC}) of .79 for the dynamic scale. The concurrent validity of the VRS:SO as a risk measure was supported, as VRS:SO static scale, pre- and post-treatment dynamic scale, and pre- and post-treatment total scale scores all had significant positive correlations with the Static-99.

The predictive accuracy of the VRS:SO for sexual recidivism was also supported by Olver et al.'s (2007) results: the Static scale, pre- and post-treatment Dynamic scale, and pre- and post-treatment Total scale scores were all significantly (p

< .001) correlated with sexual recidivism over an average 10 year follow up. Correlations ranged from .23 to .36, and AUCs from .66 to .74, comparable to the predictive validity of the Static-99, which had a significant correlation of .21 with sexual recidivism in their sample (AUC = .63). Pre- and post-treatment scores for the three dynamic factors were also all significantly correlated with sexual recidivism at follow-up. The VRS:SO Dynamic scale demonstrated significant incremental validity in the prediction of sexual recidivism over the VRS:SO Static scale, and over the Static-99. Predictive validity was also found for four interpretive risk categories of the VRS:SO Total scores, as sexual recidivism rates increased linearly across the categories from Low to High risk, with significant differences between the groups.

Although Olver et al.'s (2007) results are promising, it is important to determine whether the VRS:SO performs at a comparable level when applied to a different sample by researchers not involved with the development of the original instrument. Thus the purpose of the present study was to conduct an independent validation of the VRS-SO. The major question was whether the VRS-SO would show comparable reliability and validity for predicting sexual recidivism, and a similar factor structure as reported by Olver et al. (2007). The present study focused exclusively on the pre- and post-treatment VRS:SO scores; a comparative study of the VRS:SO change scores and other measures of treatment change is given elsewhere (Beggs & Grace, in preparation). Our sample consisted of 218 men who were convicted of a sexual offense against a child and completed the Kia Marama program between 1993 and 2000, a prison-based treatment program for child molesters in New Zealand. The VRS:SO was rated retrospectively from file review. We were also able to compare VRS:SO scores with measures of dynamic risk factors obtained from a

self-report psychometric battery, which participants completed before treatment (Allan, Grace, Rutherford, & Hudson, 2007).

1.3 Method

1.3.1 Participants

Participants were 218 adult males who completed a treatment program for sexual offenders against children while incarcerated at the Kia Marama Special Treatment Unit in Rolleston, New Zealand between 1993 and 2000. The men were aged between 18 and 74 years, with an average age of 41.1 (SD = 11.9). The majority (77.1%) were of New Zealand European ethnicity, 20.6% were New Zealand Maori, while the remaining 2.3% were from other ethnicities, including Pacific Islanders. Participants were classified according to their victim type. Ninety-five (56.4%) of the men were incest offenders, whose victims came exclusively from within their own family. The remaining 123 (43.6%) were classified as extrafamilial offenders, some or all of whose victims were unrelated to them.

1.3.2 Measures and Data Collection Procedures

1.3.2.1 VRS:SO.

The VRS:SO is a dynamic actuarial instrument for assessing pre- and post-treatment risk for sexual offenders. It consists of a 7-item static scale, and a 17-item dynamic scale. The VRS:SO static items are: Age at time of release; Age at first sexual offense; Sex offender type (incest/rapist/child molester/mixed); Prior sexual offenses; Unrelated victims; Number and gender of victims; and Prior sentencing

dates. The dynamic scale includes three factors – Sexual Deviancy, Criminality, and Treatment Responsivity (the dynamic items are listed below). VRS:SO ratings yield several scale component scores: VRS:SO Static; pre-treatment Dynamic and Total scores (i.e., Static plus pre-treatment Dynamic); post-treatment Dynamic and Total scores, and pre- and post-treatment scores on each of the three Dynamic factors.

VRS:SO Total scores can also be translated into four risk categories: Low (score of 0-20); Moderate-Low (21-30); Moderate-High (31-40); and High (41-72).

According to VRS:SO scoring protocols (Wong, Olver, Nicholaichuk, & Gordon, 2006) each dynamic item is given a rating of 0-3 prior to treatment, with higher scores being indicative of more risk. A motivational rating is also given for each item pre-treatment, based on the Stages of Change Model (Prochaska et al., 1992). For each item, the individual is assessed in terms of five stages in the model – Precontemplation, Contemplation, Preparation, Action, or Maintenance. Post-treatment VRS:SO ratings are obtained by re-assessing the stage of change for the individual on each dynamic item, and adjusting their score depending on evidence of motivational progression during treatment. For each item, the post-treatment score is the pre-treatment score minus 0.5 times the number of stages in the transtheoretical model that were progressed as a result of treatment. For example, a reduction of 0.5 is scored for progressing from Preparation to Action (one stage), or 1.0 for progressing from Contemplation to Action (two stages). An exception is that no reduction is given for progression from Precontemplation to Contemplation because of the absence of any behavioral change.

VRS:SO scores were rated retrospectively from reports and other information on file by two raters who were blind to recidivism outcomes and Static-99 scores. The structure of the Kia Marama reports was well suited to the VRS:SO rating protocols,

encompassing a description of the offender and his risk factors based on extensive assessment prior to treatment, followed by a detailed description of treatment progress and overall attitude to treatment. Twenty-three cases (10.6% of the sample) were rated independently by both raters, to allow for an analysis of the inter-rater reliability of the dynamic items.

1.3.2.2 Static-99.

The Static-99 (Hanson & Thornton, 1999) is a well-validated and widely used actuarial measure of risk among sex offenders. It consists of 10 static items which are rated on a 0-1 scale (except one item, Prior Sex Offenses, which is rated on a 0-3 scale) resulting in a maximum total score of 12, with higher scores reflecting more risk. Total Static-99 scores can be translated into risk categories, labeled Low (total score of 0-1), Medium-Low (2-3), Medium-High (4-5), and High (6 plus). Static-99 scores had previously been rated according to file information for all cases in the present study.

1.3.2.3 Recidivism.

Criminal history information was obtained from the computer database maintained by the New Zealand Department of Corrections. Convictions for sexual, violent, or general offenses that occurred post-release were noted. Sexual recidivism was defined according to the Static-99 scoring criteria for Category "A" offenses (Harris, Phenix, Hanson, & Thornton, 2003), that is, an offense with an identifiable victim (e.g., incest, sexual assault, exhibitionism). Category "B" offenses (i.e., no identifiable victim) were excluded, except for possession of child pornography. Violent recidivism was recorded when the offender had been convicted for a non-

sexual offense against a person (e.g., assault, robbery, kidnapping). General recidivism was recorded for offenses that were neither sexual nor violent (e.g., possession of cannabis). The time at large prior to each reconviction, or to the end of the follow up period, was calculated for each offender. The average follow-up time was 4.5 years (ranging between 1 and 7 years).

1.3.2.4 Psychometric measures of dynamic risk factors.

Offenders completed a psychometric battery as part of the initial assessment phase of the program. Allan et al. (2007) analyzed these data for a larger group of Kia Marama completers (N = 495) which included the present sample as a subgroup. Results of a factor analysis suggested that four factors - Social Inadequacy, Sexual Interests, Anger/Hostility, and Pro-Offending Attitudes – provided a good account of individual differences in the battery. This battery included the following tests (described in greater detail in Allan et al., 2007; and Hudson et al., 2002):

Social Inadequacy (F1) – Social Self-Esteem Inventory (SSEI; Lawson, Marshall, & McGrath, 1979); Assertion Inventory – Response Probability subscale (AI-RP; Gambrill & Richey, 1975); Fear of Intimacy Scale (FIS; Descutner & Thelen, 1991); Revised UCLA Loneliness Scale (UCLS; Russell, Peplau, & Cutrona, 1980); Hostility Towards Women scale (HTW; Check, 1985); Adult Nowicki-Strickland Internal-External Control Scale (ANSIE; Nowicki & Duke, 1983); Beck Depression Inventory (BDI) - versions I (pre-1997 participants; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and II (post-1997; Beck, Steer, & Brown, 1996); and State-Trait Anxiety Inventory (STAI; Spielberger, 1983), which includes subscales for state anxiety and trait anxiety; and the Supression subscale of the State-Trait Anger Expression Inventory (STAXI; Spielberger, 1988).

Sexual Interests (F2) – Wilson Sexual Fantasy Questionnaire, which measures the frequency of a variety of sexual fantasies and provides subscale scores for Intimate, Exploratory, Impersonal, and Sado-masochistic themes (WSFQ; Wilson, 1978).

Anger/Hostility (F3) – State-Trait Anger Expression Inventory (four remaining subscales): State Anger, Trait Anger, Anger Expression, and Anger Control.

Pro-offending Attitudes (F4) – Abel-Becker Cognitions Scale, a measure of distorted attitudes and beliefs about sexual offending against children (ABCS; Abel et al., 1989); Rape Myth Acceptance Scale (RMAS; Burt, 1980); and Adult Nowicki-Strickland Internal-External Control Scale (this test loaded onto both F1 and F4).

Factor scores were calculated for individual cases by averaging the standardized scores for tests that loaded on each factor. Resulting scores for each risk factor were positively correlated with sexual recidivism, and logistic regression analyses showed that Sexual Interests and Pro-Offending Attitudes, as well as an Overall Deviance score which aggregated all four risk factors, significantly predicted recidivism after controlling for Static-99 scores.

1.3.3 Data Analyses

Following data collection, analyses were conducted to examine the properties and validity of the VRS:SO and to attempt to replicate the major findings of Olver et al. (2007). We evaluated the inter-rater reliability of the dynamic items; the replicability of the factor structure; descriptive statistics including comparisons between incest and extrafamilial offenders; concurrent validity of the VRS:SO with the Static-99; and finally the predictive validity of the VRS:SO component scores with regard to recidivism, their incremental validity controlling for static risk, and the predictive validity of the VRS:SO risk categories. We also examined correlations

between the VRS:SO factor scores and measures of dynamic risk factors based on those derived from psychometric self-reports by Allan et al. (2007).

Statistical analyses were conducted with SPSS and Statistica.

1.4 Results

1.4.1 Inter-Rater Reliability

Reliability was assessed by computing intraclass correlation coefficients (r_{ICC}) for the 23 cases that were scored by two raters. The majority of dynamic items had significant single measure coefficients for both pre- and post-treatment. Average measures intraclass correlations for factor scores were all significant for both pre- and post-treatment, and ranged between $r_{ICC} = .79$ and $r_{ICC} = .95$ (average $r = .88$). Total Dynamic scores showed very good inter-rater reliability, $r_{ICC} = .90$, $p < .001$ for pre-treatment and $r_{ICC} = .92$, $p < .001$ for post-treatment. Overall, the reliability of the VRS:SO was acceptable, and comparable to levels reported by Olver et al. (2007).

1.4.2 Factor Analyses

We conducted a confirmatory factor analysis (CFA) to examine the extent to which the 3-factor structure reported by Olver et al. (2007) described the present data. We tested a model in which the items loading on each of the three factors were the same as reported by Olver et al (see Table 2). These were as follows: Sexual Deviance – *Sexually Deviant Lifestyle, Sexual Compulsivity, Offense Planning, Sexual Offending Cycle*, and *Deviant Sexual Preference*; Criminality – *Criminal Personality, Interpersonal Aggression, Substance Abuse, Community Support, Impulsivity*, and

Compliance with Community Supervision; and Treatment Responsivity – Cognitive Distortions, Insight, Release to High Risk Situations, and Treatment Compliance.

Standard criteria of a “good” fit or an “acceptable” fit (i.e., a Steiger-Lind RMSEA point estimate $< .05$ or $< .08$ respectively; Brown & Cudeck, 1993) were not met: RMSEA = .091 (90% CI = .078 - .104); ML χ^2 ($df = 90$) = 257.25.

Following this, exploratory factor analyses on the present data were conducted. We wanted to investigate whether a different factor structure might fit the data better, given that the sample consisted only of sexual offenders against children and therefore might be expected to have a different dynamic risk profile compared to Olver et al.’s (2007) sample, which included rapists. First we conducted a principal components analysis with varimax rotation on the pre-treatment scores for the 17 Dynamic items. The scree plot and eigenvalues suggested that a four factor solution was preferable. However first we extracted three factors (accounting for 44.69% of the variance), to assess whether the items would load consistently onto the same factors Olver et al. reported. We found some consistency – of the 15 items that loaded onto a factor in their study, 11 loaded onto the same factor in our analysis. Three items loading onto the Criminality factor in Olver et al. instead loaded onto Treatment Responsivity, and one item no longer loaded in our three-factor solution (Sexual Offending Cycle). A CFA for this solution indicated a lack of acceptable fit, similar to results for Olver et al.’s factor solution: RMSEA = .094 (90% CI = .080 - .107); ML χ^2 ($df = 90$) = 239.16.

Principal axis factoring with varimax rotation was then used to extract four factors as indicated by the scree plot and eigenvalues, which accounted for 53.1% of the total variance. The resulting factor loadings are shown in Table 2. Applying a loading cutoff of .34, 15 out of the 17 dynamic items loaded onto a single factor. The

remaining two items loaded onto two different factors; these were assigned to the factor for which the loading was highest (factor assignments are shown in bold in Table 2).

Table 2
Four-Factor Loading Matrix for VRS:SO Dynamic Items

Dynamic Item	Original Factor ^a	<i>Sexual Deviance</i>	<i>Treatment Responsivity</i>	<i>Criminality</i>	<i>Self-Management</i>
D1 Sexually deviant lifestyle	SD	.907	.081	.052	.041
D2 Sexual compulsivity	SD	.438	-.110	.408	.030
D3 Offense planning	SD	.630	-.017	.003	.125
D4 Criminal personality	C	.285	.236	.499	-.113
D5 Cognitive distortions	TR	.156	.716	.018	-.026
D6 Interpersonal aggression	C	.050	.033	.618	.185
D7 Emotional control	*	.143	-.041	.086	.591
D8 Insight	TR	-.065	.721	.021	-.003
D9 Substance abuse	C	.005	-.066	.281	.346
D10 Community support	C	.016	.388	.194	-.110
D11 Release to high risk situations	TR	-.030	.492	-.017	.035
D12 Sexual offending cycle	SD	.154	.035	.009	.704
D13 Impulsivity	C	.057	.067	.574	.206
D14 Compliance with community supervision	C	-.003	.320	.531	-.066
D15 Treatment compliance	TR	.018	.562	.352	.025
D16 Deviant sexual preference	SD	.549	-.022	.160	.050
D17 Intimacy deficits	*	.432	.051	-.004	.103

^aOriginal factors based on exploratory factor analysis reported by Olver, Wong, Nicholaichuk, & Gordon (2007). SD = Sexual Deviancy; C = Criminality; TR = Treatment Responsivity; * = item did not load

The resulting four-factor structure was similar to Olver et al.'s (2007) solution, which is also listed in Table 2 to facilitate comparison. Three of the four factors were similar to the original Sexual Deviance, Treatment Responsivity, and Criminality factors reported by Olver et al. The fourth factor included three items: Emotional Control, Substance Abuse, and Sexual Offending Cycle. This factor was labelled "Self-Management", due to the relevance of self- and emotion-regulation deficits as a commonality in the scoring criteria for these items (cf. Thornton, 2002). Comparisons between the two models indicate few differences. Two items were non-loading in

Olver et al.'s model; 12 of the 15 remaining items loaded onto the congruent factor in our four-factor model that they did in Olver et al.'s three-factor model. A confirmatory factor analysis verified that the four-factor model was an acceptable fit for these data (Brown & Cudeck, 1993), Steiger-Lind RMSEA = .076 (90% CI = .064 - .088); ML χ^2 ($df = 119$) = 280.00.

1.4.3 Descriptive Statistics and Offender Group Comparisons

Table 3 presents the means and standard deviations for both the total sample and incest and extrafamilial offenders separately for Static-99 scores, VRS:SO Static scores, and pre- and post-treatment VRS:SO Dynamic scores, Total scores, and factor scores. Note that these and all subsequent analyses are based on the dynamic risk factors as defined by Olver et al. (2007). Results of independent sample *t* tests

Table 3
Means and Standard Deviations of Static-99 and VRS:SO Scale Scores for Incest Offenders, Extrafamilial Offenders, and the Total Sample

	Incest offenders N=123 M (SD)	Extrafamilial offenders N=95 M (SD)	Total sample N=218 M (SD)
Static-99	1.3 (1.5)	3.4 (2.0) ^{***}	2.2 (2.0)
VRS:SO Static	5.7 (4.1)	10.0 (4.3) ^{***}	7.6 (4.7)
Dynamic (pre)	20.2 (5.9)	23.2 (5.5) ^{***}	21.5 (5.9)
Dynamic (post)	15.8 (6.5)	18.6 (6.3) ^{**}	17.0 (6.5)
Total (pre)	25.9 (8.4)	33.2 (8.5) ^{***}	29.1 (9.2)
Total (post)	21.5 (8.8)	28.5 (9.1) ^{***}	24.5 (9.6)
F1 Sexual deviance (pre)	8.2 (2.4)	10.3 (2.7) ^{***}	9.1 (2.7)
F1 Sexual deviance (post)	6.0 (2.5)	8.0 (2.8) ^{***}	6.9 (2.8)
F2 Criminality (pre)	4.0 (2.9)	4.3 (3.2)	4.1 (3.0)
F2 Criminality (post)	3.8 (2.8)	4.0 (3.0)	3.9 (2.9)
F3 Treatment responsivity (pre)	4.2 (2.2)	4.4 (2.1)	4.3 (2.1)
F3 Treatment responsivity (post)	3.1 (2.3)	3.3 (2.4)	3.2 (2.3)

^{**} Significantly higher score than incest offenders, $p < .01$, ^{***} $p < .001$

Statistics from independent samples *t*-tests: Static-99: $t(170.45) = 8.81$; VRS:SO Static: $t(196.58) = 7.42$; Dynamic pre: $t(209.07) = 3.85$; Dynamic post: $t(204.52) = 3.21$; Total pre: $t(201.91) = 6.28$; Total post: $t(199.43) = 5.78$; F1 pre: $t(187.42) = 5.94$; F1 post: $t(187.78) = 5.43$; F2 pre: $t(190.73) = 0.52$, *ns.*; F2 post: $t(193.13) = 0.48$, *ns.*; F3 pre: $t(208.82) = 0.70$, *ns.*; F3 post: $t(200.92) = 0.48$, *ns.*

comparing the incest and extrafamilial mean scores are also listed. As Table 3 shows, extrafamilial offenders had significantly higher risk scores than incest offenders on all measures except for Criminality and Treatment Responsivity. A similar pattern was also reported by Olver et al. (2007), who found that extrafamilial child molesters had significantly higher pre-treatment risk scores than incest offenders on all scales except Criminality and Treatment Responsivity (note that their sample also included rapists and offenders with both adult and child victims).

Overall, the mean risk scores in the present study were somewhat lower than those reported by Olver et al. (2007). For example, the average VRS:SO post-treatment Total score in the present study was 24.5 ($SD = 9.6$), compared to 32.4 ($SD = 9.9$) in Olver et al. The average Static-99 score in our sample was 2.2 ($SD = 2.0$), which falls in the Moderate-Low risk category, compared to the average of 4.4 ($SD = 2.0$) reported by Olver et al., which falls in the Moderate-High risk category (Hanson & Thornton, 1999). Thus, the present sample represented an overall lower risk profile compared to that studied by Olver et al. The exception is that mean scores on the Sexual Deviance factor were higher for our sample (pre-treatment $M = 9.1$, $SD = 2.7$) compared to Olver et al. (pre-treatment $M = 6.3$, $SD = 4.0$). An explanation for this difference is that the current sample consisted entirely of child molesters, who typically display somewhat greater sexual deviance than those who offend exclusively against adults (e.g., Yates & Kingston, 2006); indeed this pattern was found in Olver et al.

1.4.4 Concurrent Validity

The concurrent validity of the VRS:SO was assessed by computing correlations with the Static-99. Correlations between the Static-99 and VRS:SO scores

were positive: $r = .81$ with the VRS:SO Static scale; $r = .53$ with pre-treatment Dynamic; $r = .48$ with post-treatment Dynamic; $r = .76$ with pre-treatment Total; and $r = .73$ with post-treatment Total (all p 's $< .001$). Additionally, the Static and Dynamic components of the VRS:SO were correlated with each other: $r = .49$, $p < .001$ pre-treatment, and $r = .43$, $p < .001$ post-treatment. These correlations are comparable to those reported by Olver et al. (2007).

1.4.5 Predictive Validity

During the follow-up period ($M = 4.5$ years), 7.3% of all cases received convictions for a new sexual offense, 8.3% for a new violent offense, and 11.5% for a new general offense (non-sexual and non-violent). Overall, recidivism rates were lower than those reported by Olver et al. (2007; their rates were 24.6% for sexual and 35.8% violent), consistent with the lower risk profile of the current sample and shorter follow-up. For those cases that reoffended, the average time between release and reoffense was 2.0 years ($SD = 1.8$; range 36 days to 5.9 years) for sexual, 2.3 years ($SD = 1.6$; range 33 days to 5.8 years) for violent, and 1.9 years ($SD = 1.4$; range 7 days to 4.9 years) for general recidivism.

Results of correlational and Receiver Operating Characteristic (ROC) analyses of the predictive accuracy of VRS:SO scale scores (pre- and post-treatment) and the Static-99 for sexual, violent and general recidivism are presented in Table 4. The Static-99 and VRS:SO Static scale were significantly related to sexual recidivism, and to a lesser extent general recidivism, but not violent recidivism. For the Dynamic and Total scales of the VRS:SO, correlations with sexual recidivism were significantly positive for both pre- and post-treatment scores, ranging from $r = .30$, $p < .001$ for pre-treatment Total to $r = .34$, $p < .001$ for post-treatment Dynamic. Similarly, all

Table 4
Predictive Accuracy of the Static-99, and VRS:SO Scale Scores (Pre- and Post-Treatment) for Sexual, Violent, and General Recidivism

	<i>Sexual recidivism</i>			<i>Violent recidivism</i>			<i>General recidivism</i>		
	<i>r</i>	AUC	AUC 95% CI	<i>r</i>	AUC	AUC 95% CI	<i>r</i>	AUC	AUC 95% CI
Static-99	.21**	.68*	.53 - .83	.12	.59	.45 - .73	.16*	.63*	.52 - .74
<i>VRS:SO scales</i>									
Static	.17*	.67*	.54 - .81	.08	.56	.44 - .69	.15*	.63*	.53 - .73
Dynamic (pre)	.33***	.78***	.64 - .92	.18**	.63	.46 - .79	.25***	.70**	.56 - .83
Dynamic (post)	.34***	.80***	.68 - .92	.16*	.61	.45 - .77	.23***	.68**	.55 - .81
Total (pre)	.30***	.77***	.63 - .91	.16*	.61	.46 - .77	.24***	.70**	.58 - .82
Total (post)	.32***	.79***	.65 - .92	.15*	.60	.44 - .75	.23**	.69**	.57 - .81
F1 Sexual deviance (pre)	.22**	.74**	.59 - .88	.08	.58	.45 - .72	.15*	.64*	.53 - .75
F1 Sexual deviance (post)	.28***	.78***	.64 - .91	.07	.57	.42 - .71	.16*	.64*	.51 - .77
F2 Criminality (pre)	.24***	.70**	.56 - .84	.26***	.71**	.57 - .84	.37***	.77***	.67 - .88
F2 Criminality (post)	.24***	.70**	.57 - .84	.23***	.72**	.58 - .85	.37***	.77***	.66 - .88
F3 Treatment responsivity (pre)	.24***	.77***	.69 - .86	.04	.56	.42 - .70	.01	.53	.41 - .64
F3 Treatment responsivity (post)	.25***	.78***	.69 - .87	.02	.52	.37 - .67	.00	.51	.39 - .63

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

VRS:SO factor scores significantly predicted sexual recidivism both pre- and post-treatment, ranging from $r = .22, p < .01$ for pre-treatment Sexual Deviance to $r = .28, p < .001$ for post-treatment Sexual Deviance.

Pre- and post-treatment VRS:SO Dynamic and Total scales and the Sexual Deviance factor were also significantly related to general recidivism (non-sexual and non-violent), although correlations and Area Under the Curve (AUC) values were generally lower than those for sexual recidivism. Criminality factor scores were more strongly correlated with general recidivism than sexual, $r = .37, p < .001$, for both pre- and post-treatment, and also significantly predicted violent recidivism. By contrast, Treatment Responsivity factor scores were not significantly related to general or violent recidivism. The VRS:SO Dynamic and Total scales were more weakly related to violent recidivism, with correlations but not AUCs reaching significance (average $r = .16$). Sexual Deviance and Treatment Responsivity factor scores were not predictive of violent recidivism. Overall, these results are generally comparable to those reported by Olver et al. (2007), although in their data VRS:SO Static scores were a stronger predictor of sexual recidivism compared to the present sample, whereas Dynamic and Factor scores were somewhat weaker predictors.

1.4.6 Risk Categories

The sample was divided into four groups based on the post-treatment VRS:SO Total scores according to the risk category cutoffs suggested by Olver et al. (2007). These categories were: Low Risk (VRS:SO Total score 0-20, $n = 79$), Moderate-Low Risk (21-30, $n = 81$), Moderate-High Risk (31-40, $n = 41$), and High Risk (41-72, $n = 17$). Rates of sexual recidivism were significantly different across the risk categories,

$\chi^2(3, N = 218) = 33.23, p < .001$. As Figure 1 shows, the percentage reconvicted of a sexual offense increased monotonically with risk category.

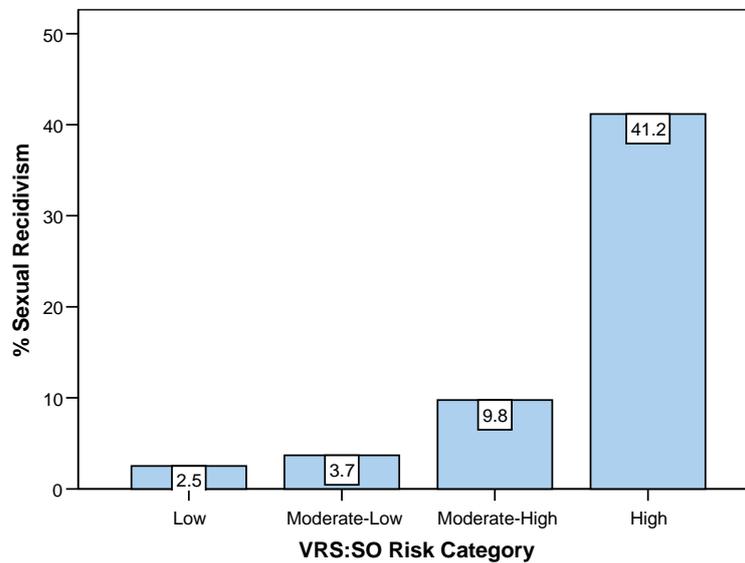


Figure 1. Recidivism base rates for risk categories derived from Total VRS:SO scores.

A Kaplan-Meier survival analysis was also conducted to compare the survival rates of the four groups (Figure 2). Pairwise comparisons (Generalized Wilcoxon, $df = 1$) showed that the High Risk group had a significantly higher rate of recidivism compared to all other groups: $\chi^2 = 11.22, p < .01$ with Moderate-High; $\chi^2 = 24.21, p < .001$ with Moderate-Low; and $\chi^2 = 37.67, p < .001$ with the Low Risk group. Additionally, the Moderate-High group reoffended at a significantly higher rate than the Low Risk group, $\chi^2 = 7.47, p < .01$.

These findings support the use of the VRS:SO risk categories to differentiate according to the likelihood of reoffending. However, it should be noted that although recidivism percentages for the two extreme categories were comparable to those in Olver et al.'s sample after a five-year follow-up, the recidivism rates in the two Moderate categories were lower in the present sample. This may suggest that although

the High Risk cutoff (VRS:SO Total score > 42) has good discriminative ability among child sexual offenders, risk increments may be lower between the other categories in comparison to samples that include rapists.

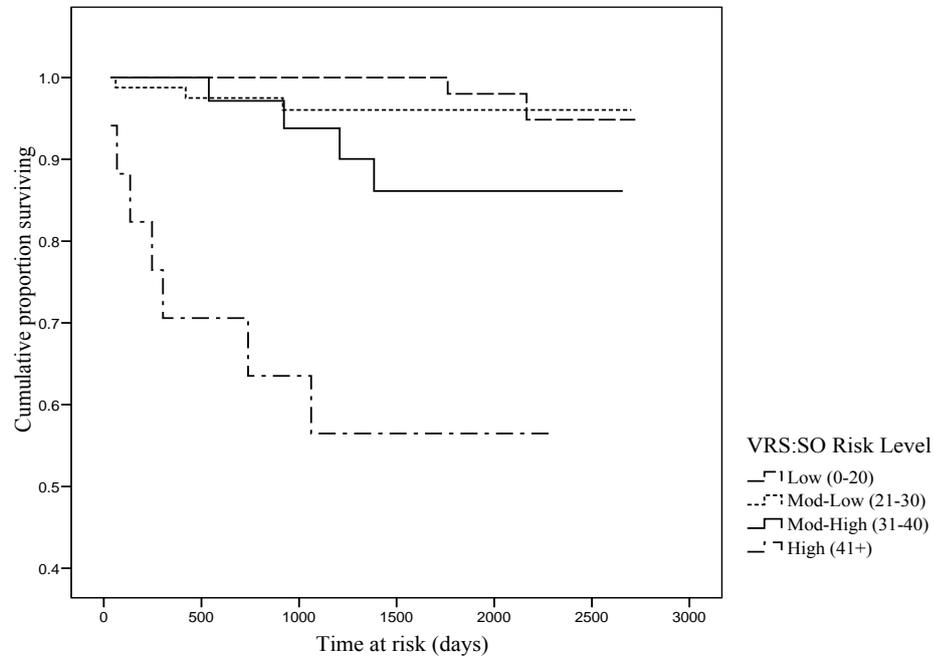


Figure 2. Kaplan-Meier survival plot showing cumulative sexual recidivism failure rates as a function of VRS:SO risk level.

1.4.7 Contributions of Dynamic Risk Controlling for Static Risk

Hierarchical multiple regressions were carried out to examine the incremental predictive validity of the VRS:SO post-treatment Dynamic scale beyond that predicted by static risk measures. Results showed that post-treatment VRS:SO Dynamic scores predicted significant additional variance with regard to sexual recidivism after controlling for the VRS:SO Static scale: Static $\beta = .03$, *ns*, Dynamic $\beta = .33$, $p < .001$, $R = .34$, $R^2 \text{ change} = .09$, $F(1, 215) = 21.63$, $p < .001$; and after controlling for the Static-99: Static-99 $\beta = .06$, *ns*, Dynamic $\beta = .31$, $p < .001$, $R = .35$,

R^2 change = .08, $F(1, 215) = 18.29$, $p < .001$. Similar results were obtained taking time at risk into account using Cox regression. The post-treatment Dynamic score was a significant predictor of sexual recidivism after controlling for VRS-SO Static, $\chi^2 = 19.91$, $p < .001$, Static Wald (1) = .05, *ns*, Exp(B) = 1.01 and Dynamic Wald (1) = 17.00, $p < .001$, Exp(B) = 1.19; and after controlling for the Static-99, $\chi^2 = 17.13$, $p < .001$, Static-99 Wald (1) = .05, *ns*, Exp(B) = 1.03 and VRS:SO Dynamic Wald (1) = 15.71, $p < .001$, Exp(B) = 1.19.

1.4.8 VRS:SO and Psychometric Measures of Dynamic Risk

Finally, we examined correlations between pre-treatment VRS-SO dynamic factor scores and the factors computed by Allan et al. (2007) based on the pre-treatment psychometric battery scores – Social Inadequacy, Sexual Interests, Anger/Hostility, and Pro-Offending Attitudes. These correlations are shown in Table 5, and provide some evidence for the convergent validity of the VRS-SO and psychometric factor scores as dynamic risk measures. The first VRS-SO dynamic factor, Sexual Deviance (F1) was positively correlated with Sexual Interests, $r = .32$, $p < .001$, but not with the other three psychometric factor scores. Anger/Hostility was positively correlated with Criminality (F2), $r = .18$, $p < .01$, but not the other two VRS-SO dynamic factors. Treatment Responsivity (F3) was positively correlated with Pro-Offending Attitudes, $r = .34$, $p < .001$. The Overall Deviance score was correlated with the Dynamic Total score, $r = .34$, $p < .001$. The general pattern of correlations was supportive of the view that the VRS-SO and psychometric self-reports were measuring similar dynamic risk factors.

Table 5 also provides correlations between the psychometric factor scores and sexual recidivism. Only the Overall Deviance score was significantly related to

recidivism, $r = .15, p < .05$. The correlations are generally lower than those reported by Allan et al. (2007), who found that significant correlations were obtained for each factor score and sexual recidivism¹. These correlations were likely attenuated because of the lower recidivism base rate in the current sample (due to the shorter follow-up time). In any case, comparison with Table 4 demonstrates that for the present data, the VRS-SO dynamic scores were more strongly predictive of sexual recidivism than factor scores based on psychometric self-reports.

Table 5
Correlations Between Factor Scores Derived From Psychometric Self Reports by Allan et al. (2007) and VRS-SO Dynamic Risk Scores and Sexual Recidivism

<i>Deviance dimension</i>	VRS:SO Factor				<i>Sexual Recidivism</i>
	<i>Sexual Deviance (F1)</i>	<i>Criminality (F2)</i>	<i>Treatment Responsivity (F3)</i>	<i>VRS:SO Dynamic Total</i>	
Social Inadequacy	.10	.02	.18**	.17*	.03
Sexual Interests	.32***	.20**	-.05	.26***	.10
Anger/Hostility	.09	.18**	.07	.19**	.13
Pro-Offending Attitudes	.05	.19**	.34***	.26***	.12
Overall Deviance	.22**	.23**	.24***	.34***	.15*

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

1.5 Discussion

The purpose of this study was to evaluate the psychometric properties and predictive validity of the Violence Risk Scale: Sexual Offender Version (VRS:SO) for sexual recidivism. The VRS:SO is a recently-developed risk instrument for sex

¹ Correlations between psychometric factor scores and sexual recidivism reported by Allan et al. (2007) were as follows: Social Inadequacy, $r = .12, p < .05$; Sexual Interests, $r = .21, p < .001$; Anger/Hostility, $r = .12, p < .05$; Pro-Offending Attitudes, $r = .22, p < .001$; Overall Deviance, $r = .28, p < .001$.

offenders designed to integrate assessment with treatment by including dynamic items as well as static. Scores on the Dynamic scale are intended to identify treatment targets, and the measure includes protocols for measuring change on these across treatment in addition to estimating risk (Olver et al., 2007). Our results supported the initial validation work by the VRS:SO developers: we found good inter-rater reliability and evidence of concurrent validity of the measure. Our findings also indicated good predictive validity with regard to sexual recidivism of the VRS:SO scale scores and risk categories. Thus our study extends the validation of the VRS:SO to a different cultural and geographical context, and to a lower-risk sample consisting exclusively of child molesters.

Results of exploratory and confirmatory factor analyses suggested that a four-factor solution provided a better fit to the present data than the three factors proposed by Olver et al. (2007). The emergence of the fourth factor, Self Management, in the present sample of medium-low risk child sexual offenders (with rapists excluded) is supported by theory and empirical findings (Thornton, 2002). The items loading onto this factor in our analysis were Emotional Control, Sexual Offending Cycle, and Substance Abuse (with a relatively low loading weight). Descriptions of these items in the VRS:SO coding manual (Wong et al., 2003) refer to sexual offending occurring following some situational, interpersonal, or personal trigger, strong emotions which are poorly controlled, or in some cases substance intoxication. This sort of offense cycle, characterized by poor self-regulation (under-control or misregulation), are features of what has been described as avoidance-goal offense pathways, in which the individual desires to avoid sexually offending, but either lacks the necessary skills or employs ineffective strategies to achieve this goal (Ward & Hudson, 1998). Although offense pathways vary between offender types and for individual offenders at

different times, a recent study found that offenders classified as following an avoidant pathway were much more likely (80%) to be child molesters than rapists or mixed offenders, and that lower risk offenders as measured by the Static-99 were also more likely to fall within the avoidance-goal pathways rather than approach (Yates & Kingston, 2006). These findings may help to explain the different distribution of scores on dynamic risk items and the emergence of the Self-Management factor in this sample, given that it consists entirely of child molesters and with an overall medium-low risk profile. Consistent with this view, Olver et al.'s (2007) three-factor structure has been replicated on a separate Canadian sample of high-risk sex offenders including rapists and child molesters (Kingston, 2003).

VRS:SO scores correlated moderately with the Static-99 in this study and in Olver et al. (2007), supporting its concurrent validity as a risk instrument. Additionally, the significant positive relationship between VRS:SO Dynamic scores and Allan et al.'s (2007) Overall Deviance and factor scores derived from a self-report psychometric battery suggest the convergent validity of the two measures as methods of assessing dynamic risk among sex offenders. The assessment of dynamic factors among sex offenders and optimal methods of incorporating these into risk assessments is a relatively new area of research. The VRS:SO and Allan et al.'s Deviance framework represent two different approaches, one based on structured therapist judgment and the other on psychometric self-report. Therapist judgements hold the advantage of avoiding the potential social desirability bias of self-report scales. The VRS:SO scores were better predictors of sexual recidivism than the Deviance scores derived from psychometric tests, suggesting that structured therapist judgment may be a more effective methodology for risk assessment than self-reports.

Overall, the VRS:SO showed good predictive validity in the present study, with pre-treatment and post-treatment Dynamic and Total scales, and pre- and post-treatment scores on all three Dynamic factors being significantly related to sexual recidivism, with AUCs in the range of .74 to .80. Incremental increases in recidivism were found across the VRS:SO Total score categories (Low, Moderate-Low, Moderate-High, High), with the High risk category being particularly discriminatory. These results suggest that combining the two Moderate categories may improve the predictive validity among samples resembling the present one. The VRS:SO Static scale (AUC = .67) was also significantly related to sexual recidivism, however importantly, the Dynamic scale was a significant predictor of sexual recidivism after controlling for static scales. This finding adds to growing evidence (e.g., Allan et al., 2007; Hanson & Harris, 2000; Olver et al., 2007) that dynamic factors can make independent contributions to risk predictions among sex offenders beyond that predicted by static factors. In fact, in this study neither the VRS:SO static scale nor the Static-99 contributed significant variance after the Dynamic scale was controlled for. This superior predictive ability of dynamic (changeable) factors over static highlights the potential for efficacious treatment that targets these variables to have a significant impact on recidivism rates. Ongoing research into which treatment targets (i.e., dynamic factors) are most strongly related to recidivism, and strategies to maximize and maintain treatment gains, could be important for further reducing recidivism. Recently, arguments have been made for a more individualized approach to sex offender treatment (Bickley & Beech, 2003; Looman, Dickie, & Abracen, 2005). Besides being a risk assessment tool, the VRS:SO has the additional function of identifying specific treatment targets for individual offenders based on pre-

treatment dynamic ratings, which could form the basis of a structured, personalized treatment program.

Alternatively, the predictive accuracy of the VRS:SO Dynamic scale may not reflect promising treatment targets, but markers of more enduring or static factors (Hanson & Harris, 2000). The greater predictive accuracy of the VRS:SO Dynamic scale compared to the VRS:SO Static scale may simply reflect its increased breadth and comprehensiveness (i.e., 17 items versus 7). A necessary next step to investigate these possibilities is to examine the changes offenders achieve on the dynamic items across treatment and the relationship between dynamic change and recidivism rates. Olver et al. (2007) reported that although the correlation between change scores on the VRS:SO Dynamic scale and sexual recidivism did not reach significance, $r = -.09$, $p = .10$, after controlling for risk level (VRS:SO Static and pre-treatment Dynamic scores) change scores demonstrated a significant unique relationship to sexual recidivism in the expected direction.

In summary, our study provides independent support for the reliability and validity of the VRS:SO as an assessment instrument for static and dynamic risk factors with child molesters. The measure showed good inter-rater reliability, concurrent and convergent validity, and compared favorably with other measures of static and dynamic risk in the prediction of sexual recidivism.

Study 2. Measuring Treatment Outcome for Child Molesters: A Comparative Validity Study

2.1 Abstract

An important concern for those working with sexual offenders is how best to measure the extent to which an individual has benefited from treatment, and how to incorporate this into risk assessments in a valid way. In the current study, three different methods of assessing proximal treatment outcome were utilised and compared on a sample of 218 men who received treatment at a prison-based group cognitive-behavioral program. These methods were: change on a psychometric battery administered pre- and post-treatment; change across treatment on the Violence Risk Scale: Sexual Offender Version (VRS:SO; Olver, Wong, Nicholaichuk, & Gordon, 2007); and post-treatment ratings on the Standard Goal Attainment Scaling for Sex Offenders (SGAS; Hogue, 1994). Problems with analysing raw psychometric change scores are highlighted, and various alternatives are explored. Results indicated the convergent validity of each of these methods, as well as predictive validity in terms of sexual recidivism, providing evidence for the association between change in dynamic risk as a result of treatment and reductions in recidivism. Results are further discussed in terms of treatment implications, and the comparative efficiency and predictive validity of the measures of treatment outcome.

There is now much evidence to suggest the effectiveness of current treatment programs for sex offenders at reducing recidivism, particularly those based on cognitive-behavioral principles (Hanson et al., 2002; Lösel & Schmucker, 2005). Although this is encouraging, the fact that a substantial proportion of men who complete treatment nonetheless go on to reoffend (approximately 10% according to a large scale meta-analysis by Hanson et al., 2002) suggests that not all who receive the same treatment will derive the same benefit. Thus, an important issue from a clinical perspective is how best to estimate the extent to which an offender has benefited from treatment, and how to incorporate this into risk assessments in a valid way.

One possibility would involve the measurement of changes across treatment on dynamic risk factors. Dynamic risk factors are variables that are related to recidivism risk and are thought to be amenable to change (such as intimacy deficits, cognitive distortions, and impaired self-regulation). These are in contrast to static risk factors, which are historical or offense-related, and by definition are unchangeable (e.g., prior sexual offenses, number of prior victims, and victim characteristics). Several studies have reported a significant relationship between identified dynamic risk factors and sexual recidivism, suggesting that dynamic factors have empirical validity as predictors of recidivism (e.g., Hanson & Harris, 2000; and see Hanson & Morton-Bourgon, 2005, for a review of empirically validated dynamic risk factors). More recently it has been shown that several risk frameworks consisting of dynamic factors are able to significantly add to the predictive validity of the Static-99 (Hanson & Thornton, 1999). These have been reviewed by Craissati and Beech (2003) and Allan, Grace, Rutherford, and Hudson (2007), and include the Sex Offender Need Assessment Rating (SONAR; Hanson & Harris, 2001 – now developed into two separate measures called the STABLE 2007 and the ACUTE 2007; Hanson, Harris,

Scott, & Helmus, 2007), Beech and colleagues' Deviancy classification (Beech, Friendship, Erikson, & Hanson, 2002), and Thornton's (2002) Initial Deviance Assessment (IDA). Allan et al. (2007) recently developed a framework for the psychometric assessment of dynamic risk in child molesters based on four dimensions: Social Inadequacy; Sexual Interests; Anger/Hostility; and Pro-offending Attitudes. Scores for each of these factors (calculated as an average of the standardized scores for the tests loading onto each factor) were significantly predictive of sexual recidivism, and scores for two of these factors (Sexual Interests and Pro-Offending Attitudes) as well as a combined Overall Deviance score, added significant predictive validity to the Static-99 in logistic regression analyses. These results indicate that self-report psychometric tests can be valid measures of dynamic risk, and also add to increasing evidence that the assessment of dynamic risk can increase the predictive accuracy of measures of static risk alone (Allan et al., 2007).

In addition to predictive validity, the other key defining premise of dynamic risk factors is that they are in fact dynamic, implying that they are amenable to change. However, relatively few studies have measured dynamic risk at more than one point in time and assessed whether change on these factors is associated with a reduction in risk. In one exception, Hudson, Wales, Bakker, and Ward (2002) looked at differences in the pre-treatment and post-treatment scores of 242 child molesters on a battery of psychometric tests intended to assess sexual attitudes and beliefs, emotional functioning, and interpersonal competency (the sample of this study was an earlier subsample of that used in Allan et al., 2007, and included largely the same psychometric battery). Hudson et al. (2002) found that change in a pro-social direction was generally related to reduced recidivism, however correlations were only significant for a handful of tests (assessing empathy, assertiveness, and trait anger). Paradoxically, change on some tests (measuring anger suppression and frequency of

impersonal and sado-masochistic sexual fantasy) had significant correlations with recidivism in the opposite direction to expectations, such that pro-social change was associated with increased recidivism. One possible explanation for this is that those who were most deviant prior to treatment had a greater opportunity to show improvement. This highlights a potential problem with using raw change scores to assess treatment gains.

Olver, Wong, Nicholaichuk, and Gordon (2007) developed an alternative method of assessing change in dynamic factors across treatment and incorporating these into sex offender risk estimates. Their Violence Risk Scale: Sexual Offender Version (VRS:SO) consists of a 7-item static scale and a three-factor 17-item dynamic scale. Changes on the dynamic items are measured following treatment implementing a scoring protocol based on the Transtheoretical Change Model (Prochaska, DiClemente, & Norcross, 1992). In their recent validation study of the VRS:SO, Olver et al. (2007) reported that the measure showed good inter-rater reliability, and predictive validity with regard to sexual recidivism, findings that have been replicated in an independent validation study by Beggs and Grace (2008). Olver et al. (2007) additionally reported that change on the Dynamic scale was related to sexual recidivism and contributed significant additional variance beyond the VRS:SO Static scale. They also divided their sample into “High” and “Low” risk groups according to Static-99 scores, and found that change on the Dynamic scale was predictive of recidivism only among the High risk group, suggesting that therapeutic change is more predictive among higher risk offenders. These findings require replication, however the VRS:SO appears to show much promise as a method of assessing changes in dynamic risk across treatment in a valid way that is related to decreased recidivism.

Aside from measuring change on dynamic risk factors, another approach to assessing the extent to which an offender has benefited from treatment involves the use of structured post-treatment clinical rating systems. This type of methodology has been applied in several studies in which the relationship between treatment outcome and recidivism has been investigated (Barbaree, 2005; Langton, Barbaree, Harkins, & Peacock, 2006; Looman, Abracen, Serin, & Marquis, 2005; Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005; Quinsey, Khanna, & Malcolm, 1998; Scalora & Garbin, 2003; Seager, Jellicoe, & Dhaliwal, 2004; and Seto & Barbaree, 1999). The rating systems utilized in these studies vary in terms of content and complexity, but are usually based on behavioral descriptors and rated by clinicians (or researchers from file review). However, results of studies investigating the relationship between such rating systems and recidivism have been mixed. Several studies (including Quinsey et al., 1998, in which overall treatment gains were rated on a four-point scale from “Poor” to “Very Good”; and Seager et al., 2004, in which offenders were given pass/fail ratings on five components of treatment) reported no relationship between treatment outcome and reoffending (Quinsey et al., 1998, also reported that the treatment program they were evaluating appeared to be ineffective at reducing recidivism). Seto and Barbaree (1999) scored their sample on “Treatment Behavior” (a summed score based on dichotomous ratings for items relating to behavior in group, motivation for treatment, and change achieved in treatment), and reported that contrary to expectations, positive ratings of treatment behavior were related to *increased* serious recidivism (sexual and/or violent). However, Barbaree (2005) subsequently analyzed results for the same sample with a longer follow-up time and more thorough recidivism data, and found that there was no relationship between treatment behavior and serious recidivism. This finding was later replicated with

regard to sexual recidivism by Langton et al. (2006) using a revised version of the rating scale.

Finally, three studies using relatively simple dichotomous rating systems according to whether offenders had reduced their risk or not (Looman et al., 2005), “got it” or “did not get it” (Marques et al., 2005); or were “successfully treated” or “unsuccessfully treated,” (Scalora & Garbin, 2003) have reported a significant relationship between positive ratings of treatment outcome and reduced sexual recidivism. Looman et al. also applied a separate measure of treatment behavior in which ratings on a 4-point scale were given for three treatment targets and then summed; this scale was unrelated to recidivism.

Overall, these findings indicate that structured rating systems of treatment outcome have potential to assist risk assessments among treated sex offenders. However, the variability of findings regarding the relationship between these measures and recidivism (including within-study variability when more than one rating system was implemented, as in Looman et al., 2005) suggests the importance of ensuring the validity of a measure prior to applying it clinically. One measure that shows promise in this respect is Standard Goal Attainment Scaling for Sex Offenders (SGAS; developed by Hogue, 1994). This measure involves rating offenders on a 5-point scale according to the extent to which they have attained the goals of treatment (e.g., show empathy for their victims, accept personal responsibility for their offending, and recognize cognitive distortions). In an initial validation study of the SGAS, Hogue (1994) found evidence for the inter-rater reliability, face validity, and convergent validity of the measure, and it has since been implemented in studies to track treatment progress (Barrett, Wilson, & Long, 2003; Stirpe, Wilson, & Long, 2001), however it is yet to be validated in terms of predictive ability for sexual recidivism.

In the current study, three different methods of assessing gains made in treatment were applied on a sample of child molesters who completed a group cognitive-behavioral program while incarcerated. These were: change on a four-factor psychometric risk assessment framework (developed by Allan et al., 2007); change on the three-factor Dynamic scale of the VRS:SO; and post-treatment ratings of the attainment of treatment goals using a modified version of the SGAS. The specific research questions included what each method would reveal in terms of the extent to which the sample had benefited from the treatment program, and whether there would be agreement between the methods (indicative of convergent validity). We were also interested in the relationship between proximal treatment outcome and recidivism, and the predictive validity of the measures of treatment outcome relative to each other and to a measure of static risk (Static-99).

2.3 Method

2.3.1 Participants

The participants in this study were adult males who had completed a 32-week prison-based group treatment program for men who have sexually offended against children (Kia Marama Special Treatment Unit, Rolleston Prison, New Zealand). This program is based on cognitive-behavioral principles, with an underlying relapse prevention framework. Groups of ten men attend three hour therapy sessions three times per week, covering the following modules: Norm Building, Offense Chains; Arousal Reconditioning; Victim Impact and Empathy; Mood Management; Relationship Skills, Sexuality, and Relapse Prevention (for a more thorough description of this program, see Hudson, Marshall, Ward, Johnston & Jones, 1995; or Hudson, Wales, & Ward, 1998). All participants gave written consent for their file

information to be used for research and evaluation purposes when entering the program.

The total sample consisted of the first 495 men who completed the Kia Marama program for which criminal history, demographic details, psychometric test scores, and recidivism data was available (the same sample as that used in Allan et al., 2007). These men were aged between 18 and 76, with an average age of 41.0 (SD = 12.2). 80.8% of the sample was of European descent, 16.7% were New Zealand Maori, and the remaining 2.5% were of other ethnicities. The sample included approximately even numbers of incest offenders (52.3%) whose victims were exclusively from within their own family, and extrafamilial offenders (47.7%) who were unrelated to some or all of their victims. Preliminary analyses indicated no significant differences between incest and extrafamilial offenders on measures of treatment outcome or recidivism, so only the results for the entire sample are presented below. A subsample of 223 were rated on the SGAS and VRS:SO measures which required a labor-intensive file review; VRS:SO scores could not be obtained for 5 of these men due to missing file information, resulting in a sample size of $N = 218$ for analyses involving this measure.

2.3.2 Measures of Treatment Outcome

Three separate methods of assessing treatment progress were employed in this study. These were: Changes on a self-report psychometric battery administered pre- and post-treatment; Change scores on the Violence Risk Scale: Sexual Offender Version (VRS:SO), a measure encompassing 17 dynamic items assessed pre-treatment and then adjusted post-treatment according to progression through five stages of change (Wong et al., 2006); and Standard Goal Attainment Scaling (SGAS)

scores, consisting of post-treatment ratings of the level of attainment of six treatment goals (Hogue, 1994). Each of these is described in more detail below.

2.3.2.1 *Psychometric change.*

A battery of self-report psychometric tests was administered to the sample prior to treatment and again after completion of the program, assessing across four domains identified by Allan et al.'s (2007) factor analysis: Social Inadequacy, Sexual Interests, Anger/Hostility, and Pro-offending Attitudes. This battery included the following tests (described in greater detail in Allan et al., 2007; and Hudson et al., 2002):

Social Inadequacy (F1) – Social Self-Esteem Inventory (SSEI; Lawson, Marshall, & McGrath, 1979); Assertion Inventory – Response Probability subscale (AI-RP; Gambrill & Richey, 1975); Fear of Intimacy Scale (FIS; Descutner & Thelen, 1991); Revised UCLA Loneliness Scale (UCLS; Russell, Peplau, & Cutrona, 1980); Hostility Towards Women scale (HTW; Check, 1985); Adult Nowicki-Strickland Internal-External Control Scale (ANSIE; Nowicki & Duke, 1983); Beck Depression Inventory (BDI) - versions I (pre-1997 participants; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and II (post-1997; Beck, Steer, & Brown, 1996); and State-Trait Anxiety Inventory (STAI; Spielberger, 1983), which includes subscales for state anxiety and trait anxiety; and the Suppression subscale of the State-Trait Anger Expression Inventory (STAXI; Spielberger, 1988).

Sexual Interests (F2) – Wilson Sexual Fantasy Questionnaire, which measures the frequency of a variety of sexual fantasies and provides subscale scores for Intimate, Exploratory, Impersonal, and Sado-masochistic themes (WSFQ; Wilson, 1978).

Anger/Hostility (F3) – State-Trait Anger Expression Inventory (four remaining subscales): State Anger, Trait Anger, Anger Expression, and Anger Control.

Pro-offending Attitudes (F4) – Abel-Becker Cognitions Scale, a measure of distorted attitudes and beliefs about sexual offending against children (ABCS; Abel et al., 1989); Rape Myth Acceptance Scale (RMAS; Burt, 1980); and Adult Nowicki-Strickland Internal-External Control Scale (this test loaded onto both F1 and F4).

2.3.2.2 *Violence Risk Scale: Sexual Offender Version change scores.*

The Violence Risk Scale: Sexual Offender Version (VRS:SO) is a measure designed to assess the risk of reoffending among sexual offenders. It consists of a 7-item static scale, and a 17-item dynamic scale consisting of the following three factors: F1 Sexual Deviancy (Sexually deviant lifestyle, Sexual compulsivity, Offense planning, Sexual offending cycle, and Deviant sexual preference); F2 Criminality (Criminal personality, Interpersonal aggression, Substance abuse, Community support, Impulsivity, and Compliance with community supervision); and F3 Treatment Responsivity (Cognitive distortions, Insight, Release to high risk situations, and Treatment compliance); and there are also two non-loading factors (Emotional control, and Intimacy deficits).

According to VRS:SO scoring protocols (Wong et al., 2006), each of these items is rated at pre-treatment on a 4-point scale from 0-3, with higher scores reflecting increased risk. The dynamic items are also rated according to which of five stages of change the offender is at for the particular item – Precontemplation, Contemplation, Preparation, Action, or Maintenance (Prochaska et al., 1992). Post-treatment scores are obtained by re-rating the stages of change for each dynamic item, and then subtracting 0.5 points from the pre-treatment item score for each stage progressed (however no point reduction is given for progression from Precontemplation to Contemplation due to absence of observable behavior change). For example, someone with a pre-treatment risk rating of 2 for a dynamic item who

progressed from either the Precontemplation or Contemplation stage to the Action stage would receive a post-treatment score of 1 (i.e., $2 - (2 \times 0.5)$, for progressing two stages). VRS:SO Dynamic change scores were obtained by subtracting the post-treatment Dynamic score from the pre-treatment Dynamic score for each participant. Factor change scores were also obtained in this way for each of the three VRS:SO factors.

For this study the VRS:SO was rated from file information including treatment reports, case notes, and offense history documents by two independent coders who were blind as to recidivism outcome. Good inter-rater reliability was found for Dynamic scores both pre-treatment, $r_{ICC} = .90, p < .001$, and post-treatment, $r_{ICC} = .92, p < .001$. The reliability of Dynamic Change scores was also good, $r_{ICC} = .82, p < .001$.

2.3.2.3 Standard Goal Attainment Scaling.

A modified version of Hogue's (1994) Standard Goal Attainment Scaling for Sex Offenders (SGAS) was used to rate the extent to which participants had attained six goals of treatment, based on post-treatment reports on file. Attainment of each goal was rated on a five-point scale ranging from -2 to +2 (with a score of zero representing a minimally acceptable level of goal attainment), resulting in a Total SGAS score ranging from -12 to +12. The goals were: 1) Show empathy and insight into victim issues; 2) Accept personal responsibility for offending; 3) Recognize cognitive distortions; 4) Understand offense chain; 5) Identify relapse prevention concepts; and 6) Motivation to change behavior. SGAS scores were rated by two independent data coders from post-treatment reports on file written by the group therapist, with a good level of inter-rater reliability ($r_{ICC} = .88, p < .001$). The coders

were blind as to recidivism outcomes while conducting the ratings, and completed these independently of each other.

2.3.3 Static Risk Measure

The Static-99 (Hanson & Thornton, 1999) is a risk assessment instrument for sex offenders that consists of 10 static items: Young age; Ever lived with intimate partner; Index non-sexual violence; Prior non-sexual violence; Prior sex offenses; Prior sentencing dates; Any convictions for non-contact sex offenses; Any unrelated victims; Any stranger victims; and Any male victims. Each of these items is scored on a 0-1 scale, except for Prior sex offenses which is scored on a 0-3 scale. Total scores can therefore range from 0-12, with higher scores indicating a greater likelihood of future sexual offending. The Static-99 has consistently been shown to have moderate to good predictive validity for sexual recidivism in a range of evaluation studies (e.g., G. T. Harris et al., 2003), and was included in this study to allow for analyses of the relationship between treatment change and recidivism independent of static risk.

2.3.4 Recidivism

Categorical data on reconviction was obtained from the official New Zealand criminal convictions database after an average follow-up period of 5.8 years following release. Data on sexual, violent (excluding sexual violence), and general reconvictions were recorded separately.

2.4 Results

2.4.1 Psychometric Change

We employed several methods for analyzing treatment change based on self-reported psychometric measures. Table 6 shows mean pre-treatment and post-treatment scores for the sample on each of the psychometric tests (note that some tests were reverse scored so that higher scores indicate greater deviancy for every test, and that the sample size varies across measures due to changes in the test battery administered at Kia Marama over time). As a first approach to measuring treatment change, we computed the difference between pre- and post-treatment scores. These change scores were calculated such that positive values indicated change in a pro-social direction (i.e., less deviant attitudes or improved social functioning). Effect sizes are presented in the far right column of Table 6, along with the significance levels from corresponding *t* tests. Medium to large effect sizes indicating pro-social changes were obtained for many of the variables, particularly those from F1 Social Inadequacy and F4 Pro-offending Attitudes. The largest positive change was obtained for distorted cognitions regarding children and sexual behavior (ABCS), $d = 0.82$, but substantial gains were also obtained for rape myth acceptance, $d = 0.55$, internal locus of control, $d = 0.52$, assertiveness, $d = 0.58$, anger suppression, $d = 0.52$, fear of intimacy, $d = 0.60$, hostility toward women, $d = 0.53$, and depression, $d = 0.70$. Thus, the change scores suggest that moderate to substantial treatment gains were obtained in terms of reduced pro-offending attitudes and increased social adequacy.

However, these results should be interpreted cautiously, because there are clear threats to validity when making pre- vs. post-treatment comparisons. First, a self-presentation bias might be anticipated, because the men had obvious incentives to show improvement as a result of treatment (e.g., to earn positive recommendations from program staff). Moreover, it is well known that use of difference scores to measure change poses methodological difficulties (Cronbach & Furby, 1970). One problem is related to the fact that offenders differ in their pre-treatment levels of

Table 6
*Means, Standard Deviations, and N for Pre-Treatment and Post-Treatment Psychometric Variables, and Treatment Change Effect Sizes (Cohen's *d*).*

<i>Test</i>	<i>Pre-treatment</i>			<i>Post-treatment</i>			<i>Effect size^a</i>
	Mean	SD	N	Mean	SD	N	
<i>Social Inadequacy (F1)</i>							
Social Self-Esteem Inventory	112.96	27.97	474	120.81	26.38	450	0.28***
Assertion Inventory							
Response Probability	113.29	22.00	417	101.04	24.86	382	0.58***
State Trait Anger Expression Inv.							
Suppression	18.24	4.57	416	15.72	4.42	402	0.52***
Fear of Intimacy Scale	93.76	23.30	284	87.13	22.11	288	0.60***
UCLA Loneliness	46.13	10.15	383	39.97	11.06	372	0.30***
Hostility Toward Women	11.89	6.40	471	8.57	6.04	440	0.53***
A N-S Internal-External Control	15.72	6.04	473	12.53	5.83	445	0.52***
Beck Depression Inventory	15.99	9.74	454	8.96	8.54	355	0.70***
State Trait Anxiety Inventory							
State	39.98	13.67	472	33.24	12.14	455	0.47***
Trait	44.33	12.28	472	37.97	11.30	455	0.49***
<i>Sexual Interests (F2)</i>							
Wilson Sex Fantasy Questionnaire							
Exploratory	10.89	8.50	471	7.80	7.12	438	0.36***
Intimate	23.69	11.16	471	23.55	11.51	438	0.02
Impersonal	11.39	7.87	471	8.12	6.54	438	0.42***
Sado/masochistic	4.12	5.73	471	2.89	4.98	438	0.23***
<i>Anger/Hostility (F3)</i>							
State Trait Anger Expression Inv.							
State	13.60	6.28	425	11.72	4.00	401	0.29***
Trait	18.93	6.00	427	17.26	5.36	401	0.28***
Expression	15.97	4.65	425	15.48	3.78	402	0.11***
Control	22.04	5.66	425	22.85	5.54	402	0.14***
<i>Pro-offending Attitudes (F4)</i>							
Abel-Becker Cognition Scale	51.73	15.98	471	38.79	10.40	441	0.82***
Rape Myth Acceptance Scale	48.49	19.07	465	37.63	17.53	449	0.55***
A N-S Internal-External Control	15.72	6.04	473	12.53	5.83	445	0.52***

^a Effect sizes were computed as the difference between pre-treatment and post-treatment scores divided by the standard deviation of change scores, such that positive values indicate prosocial change.

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

deviancy. Because the psychometric tests have minimum scores, the maximum possible change score increases as a linear function of the pre-treatment score. As a result, more deviant or dysfunctional offenders have greater opportunities to demonstrate pro-social change. This is problematic, given the strong incentives to

show improvement, and the transparent nature of some of the psychometric tests (e.g., ABCS; WSFQ).

To test whether there was a relationship between pre-treatment and change scores, we conducted a correlational analysis. For each psychometric variable, the correlation was positive and significant such that greater treatment gains were associated with more deviant or dysfunctional pre-treatment scores; the average correlation was $r = .57$ and all were .40 or greater. Following Krause, Howard and Lutz' (1998) recommendation to explore change data at a disaggregated level, we examined scatterplots for all variables. These confirmed the strong linear relationship between pre-treatment and change scores.

To illustrate this relationship, Figure 3 shows a scatterplot for the ABCS (reverse scored, so that higher pre-treatment scores indicate greater deviance). There is a strong positive correlation between the pre-treatment and change scores, $r = .80$. Figure 3 also shows that pro-social change was reported by nearly the entire sample, as indicated by change scores greater than zero. The large circle at the left of the cluster of points represents the 14 offenders who reported minimally-deviant scores at both pre- and post-treatment. However, the most dramatic feature of Figure 3 is the ceiling effect on change scores. This occurred because a large percentage of the sample (17.8%) reported the minimum post-treatment score (i.e., least deviant); in contrast, only 4.2% reported the minimum score at pre-treatment. When the minimum post-treatment score is obtained, the change score is a linear function of the pre-treatment score; hence, the clear limit on maximum change as the pre-treatment score increases. Scatterplots for other variables showed similar results in many cases. These findings raise the possibility that there is a self-presentation or impression management bias at post-treatment, with offenders eager to show that the program had been effective for them. This bias would produce an artefactual correlation between

pre-treatment and change scores such that offenders who are more deviant pre-treatment can potentially report greater change. Of course, there is also the possibility that no such bias exists, but that higher scorers pre-treatment simply have more room to show genuine improvement.

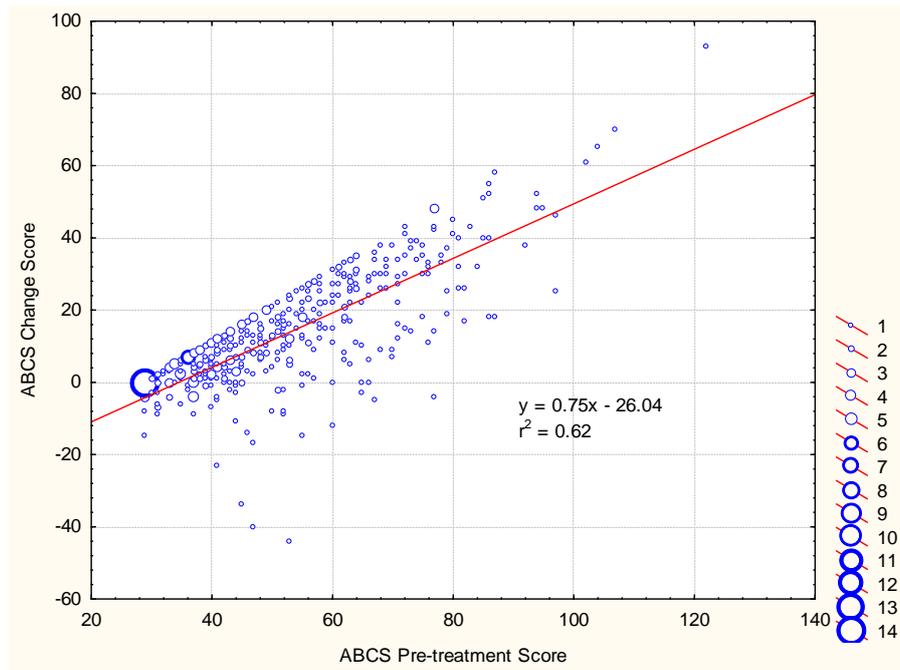


Figure 3. Change scores (with positive values indicating treatment gains) for the Abel-Becker Cognition Scale (ABCS) as a function of pre-treatment score. Different-sized circles in the legend correspond to the number of cases that reported particular combinations of pre-treatment and change scores. The regression line, equation, and variance accounted for are also shown.

Because of these difficulties with raw change scores, we pursued two additional strategies towards assessing treatment change using the psychometric self-reports. Our first strategy involved partialling out the pre-treatment scores and calculating standardized residual change scores. The second strategy was to apply criteria for clinically significant and reliable change to offenders' scores (described in greater detail below). For both methods, we then calculated average scores for measures on the different dynamic risk factors identified by Allan et al. (2007). In

addition to addressing the problems with analyzing raw difference scores, these alternative strategies involved standardization across the battery, enabling both comparisons across tests and determination of average change within each dynamic risk factor and overall.

2.4.1.1 Standardized residual change scores.

To attempt to control for the relationship between pre-treatment and change scores statistically, we regressed the raw change scores onto the pre-treatment scores. In every case, the pre-treatment scores accounted for a significant proportion of the variance in the change scores. Across measures, R^2 values ranged from .16 to .65, with an average of .34. Thus, a substantial portion of the variance in change scores was associated with variation in pre-treatment scores. To remove this variance, we calculated the residuals from the regressions (i.e., obtained change score – predicted change score), and then standardized the residuals for each variable. Analyses using average standardized residual change scores (RCZ) across the four factors identified by Allan et al. (2007) and overall are reported below.

2.4.1.2 Clinically significant and reliable change.

We next examined whether offenders demonstrated clinically significant and reliable change on each of the psychometric measures (a methodology described by Evans, Margison, & Barkham, 1998). Clinically significant change refers to whether, at post-treatment, an offender's scores on the psychometric tests are indistinguishable from normative (i.e., non-offender) samples. This is calculated by first determining a

cut-off score – a mid point between the distributions of normative¹ and non-normative scores (formula² obtained from Evans et al., 1998). For each offender and measure, we then determined if the pre- and post-treatment scores were: a) above and below the cut-off, respectively; b) below and below; c) above and above; or d) below and above. Offenders in categories a) and b) were defined as having demonstrated “clinically significant change” on the measure in question, because the post-treatment scores were below the cut-off. Category a) itself is also interesting because these offenders changed from deviant to non-deviant during treatment. However, because scores from repeated test administrations can vary by chance, it is important also to consider the reliability of the change between pre- and post-treatment scores. Reliable change refers to whether or not the offender’s level of change on a measure is beyond that which could be attributed to the measurement variability of the test itself.

Determining whether an offender’s change on a particular test is reliable requires calculating the standard error of the difference (SE_{diff} ³; Evans et al., 1998) based on the pre-treatment score distribution and the reliability of the measure¹. Change exceeding 1.96 times the SE_{diff} is unlikely to occur more than 5% of the time due to test unreliability alone, therefore an offender’s level of change on a test can be compared against this criterion to determine whether they achieved reliable change.

¹ Normative and reliability data for calculating Reliable and Clinically Significant Change were obtained from adult male samples for the following measures: ABCS – Tierney and McCabe (2001); AI-RP – Gambrill and Richey (1975); ANSIE – Nowicki & Duke (1974), S. Nowicki (personal communication, 2007); BDI – Beck, Steer, and Garbin (1988), Seggar, Lambert, and Hansen (2002); FIS – Descutner and Thelen (1991), Doi and Thelen (1993); HTW – Dean and Malamuth (1997), Malamuth, Sockloskie, Koss, and Tanaka (1991); RMAS – Burt (1980), Dean and Malamuth (1997); SSEI – Lawson, Marshall, and McGrath (1979); STAI – Spielberger (1983); STAXI – Spielberger (1988); UCLS – Russel, Peplau, and Cutrona (1980); WSFQ – Baumgartner, Scalora, and Huss (2002), Plaud and Bigwood (1997).

$$^2 \text{ CSC cutoff} = \frac{(mean_{clin} \times SD_{norm}) + (mean_{norm} \times SD_{clin})}{SD_{norm} + SD_{clin}}$$

³ $SE_{diff} = SD_1 \sqrt{2\sqrt{1-r}}$ where SD_1 is the standard deviation of the baseline observations, and r is the reliability of the measure.

Table 7 shows the percentage of the sample who achieved clinically significant change (CSC), and both reliable and clinically significant change (RCSC) on each of the psychometric measures. Note that the CSC classification refers only to whether the offenders' post-treatment scores were below the cutoff for the normative range, and therefore includes offenders whose scores were below the cutoff pre-treatment (i.e., non-deviant). The average percentages for tests within each of the four factors identified by Allan et al. (2007) are also given, as well as the overall average across all four factors. As can be seen, relatively high proportions of the sample attained CSC on the tests (the average across all tests was 65.2%). Although this seems encouraging, not all of the gains can be attributed to the treatment program. The second column from the right in Table 7 indicates the percentage of those attaining CSC post-treatment who actually had shifted from pre-treatment scores within the deviant range (labelled above as category a). Across the entire battery, this average was 36.6% of those attaining CSC, or 23.9% of the full sample. The remainder of scores were not deviant prior to treatment. Similarly, averaging across all tests, only an average of 19.4% attained a level of change that was both reliable and clinically significant (RCSC). Tests assessing Social Inadequacy (F1) and Pro-offending Attitudes (F4) had on average higher percentages of RCSC than the other factors, a similar pattern to the effect size analyses reported above.

Thus despite the medium to large effect sizes for psychometric change, and the fact that after completing treatment nearly two-thirds of men showed overall scores on the battery that were indistinguishable from non-offender norms, in the majority of cases (63.4%) the pre-treatment scores were in the normative range. Overall, about one in five of the sample (19.4%) had non-deviant psychometric scores post-treatment and had also evidenced a level of change on the tests across the program beyond that

Table 7

Percentage that Achieved Clinically Significant Change (CSC), Percentage Achieving CSC who were Deviant Pre-Treatment, and Percentage that Achieved both Reliable and Clinically Significant Change (RCSC) on Each Test

<i>Test</i>	<i>N</i>	<i>% CSC</i>	<i>% CSC deviant pre-tx (a)</i>	<i>% RCSC</i>
<i>Social Inadequacy (F1)</i>				
Social Self-Esteem Inventory	446	58.0	16.1	5.8
Assertion Inventory				
Response Probability	373	58.9	45.8	22.0
State Trait Anger Expression Inv.				
Suppression	391	61.2	51.2	21.7
Fear of Intimacy Scale	266	50.0	44.4	24.4
UCLA Loneliness	328	56.2	43.5	31.7
Hostility Toward Women	435	63.2	46.4	21.4
A N-S Internal-External Control	439	49.4	54.5	14.6
Beck Depression Inventory	342	65.6	52.8	26.9
State-Trait Anxiety Inventory				
State	450	69.5	42.1	28.4
Trait	450	52.7	50.4	23.3
F1 average	450	58.5	44.7	22.0
<i>Sexual Interests (F2)</i>				
Wilson Sex Fantasy Questionnaire				
Exploratory	431	78.3	28.3	18.3
Intimate	430	63.5	18.3	16.0
Impersonal	430	74.2	32.9	17.7
Sado-masochistic	430	77.9	21.1	11.6
F2 average	431	73.5	25.2	15.9
<i>Anger/Hostility (F3)</i>				
State-Trait Anger Expression Inv.				
State	397	82.0	23.1	15.9
Trait	399	66.3	30.8	15.3
Expression	398	55.2	30.6	5.5
Control	398	59.5	20.1	2.5
F3 average	399	65.8	26.2	9.8
<i>Pro-offending Attitudes (F4)</i>				
Abel-Becker Cognition Scale	436	83.4	49.7	42.7
Rape Myth Acceptance Scale	442	79.3	28.7	22.2
A N-S Internal-External Control	439	49.4	54.5	14.6
F4 average	442	70.7	44.3	26.5
<i>Overall average</i>	<i>450</i>	<i>65.2</i>	<i>36.6</i>	<i>19.4</i>

which could be attributable to measurement error, whereas 34.8% on average had scores in the deviant range post-treatment.

2.4.2 Measures of Treatment Outcome Based on File Review – VRS:SO and SGAS

Measures of treatment outcome based on file review or clinician ratings (such as SGAS, and change on the VRS:SO) could be expected to avoid the drawback of the possible self-presentation bias associated with self-report tests. These measures may potentially be able to provide a more accurate description of treatment outcome among this sample.

Table 8
Means and Standard Deviations of VRS:SO Change Scores ($N = 218$), and SGAS Scores ($N = 223$)

<i>Measure of Treatment Outcome</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Violence Risk Scale: Sexual Offender Version				
F1 Change – Sexual deviance	0	4.50	2.25	1.07
F2 Change – Criminality	0	2.0	0.26	0.38
F3 Change – Treatment responsiveness	0	3.0	1.11	0.61
Total VRS:SO Dynamic Change	0	8.50	4.53	1.89
Standard Goal Attainment Scaling				
Show empathy & insight into victim issues	-2	2	-.21	1.00
Accept personal responsibility	-2	2	-.12	.87
Recognize cognitive distortions	-2	2	-.06	.76
Understand offense chain	-2	2	.37	.81
Identify relapse prevention concepts	-2	2	.35	1.02
Motivation to change behavior	-2	2	.27	.99
SGAS Total	-10	10	.59	4.14

The range, means, and standard deviations of VRS:SO change scores and SGAS scores are shown in Table 8. The average point reduction achieved across treatment on the VRS:SO Dynamic scale was 4.53 ($SD = 1.89$), ranging from two men in the sample (0.9%) who did not change at all on any of the 17 dynamic items (change score of 0) to the sample maximum of 8.5 achieved by three individuals. The majority of these changes occurred on the items relating to Sexual Deviance (F1), and very little change in general occurred on the Criminality items (F2). For the SGAS, almost two thirds of the sample (63.2%) received a total score greater than zero, which as per the scoring protocols indicates at least a minimally-acceptable level of

goal attainment achieved in the program. The average Total SGAS score was 0.59 ($SD = 4.14$). Average scores for the six SGAS goals ranged from a low of -0.21 ($SD = 1.00$) for “Show empathy and insight into victim issues,” to 0.37 ($SD = 0.81$) for “Understand offense chain.”

Table 9

Correlations Between Measures of Treatment Outcome: Standardized Residual Change (RCZ), Clinically Significant Change (CSC) and Reliable and Clinically Significant Change (RCSC) on the Four-Factor Psychometric Battery; Standard Goal Attainment Scaling (SGAS) Total Scores; and Change Scores on the Violence Risk Scale: Sexual Offender Version (VRS:SO) Total Dynamic Scale and Three Factors

<i>Measure of Treatment Outcome</i>	<i>VRS:SO Change</i>			<i>Total Dynamic</i>	<i>SGAS Total</i>
	<i>F1</i>	<i>F2</i>	<i>F3</i>		
<i>Psychometric RCZ</i>					
F1 Social Inadequacy	.25***	.00	.12	.25***	.33***
F2 Sexual Interests	.00	-.03	-.07	-.04	-.05
F3 Anger/Hostility	.12	-.02	.01	.11	.13
F4 Pro-offending Attitudes	.28***	.02	.13	.27***	.28***
Overall RCZ	.22**	-.02	.07	.21**	.23***
<i>Psychometric CSC</i>					
F1 Social Inadequacy	.23**	.02	.02	.21**	.35***
F2 Sexual Interests	.01	-.07	-.06	-.02	.04
F3 Anger/Hostility	.05	-.09	-.02	.04	-.01
F4 Pro-offending Attitudes	.23**	.02	.05	.20**	.27***
Overall CSC	.19**	-.05	.00	.16*	.24***
<i>Psychometric RCSC</i>					
F1 Social Inadequacy	.22**	.04	.10	.25***	.27***
F2 Sexual Interests	-.01	.01	.00	.01	-.06
F3 Anger/Hostility	.21**	.15*	.07	.24***	.08
F4 Pro-offending Attitudes	.09	.09	.06	.06	.05
Overall RCSC	.20**	.11	.10	.24***	.14*
SGAS Total	.66***	.16*	.42***	.69***	

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

Table 9 shows the correlations between the three measures of treatment outcome utilized in this study – psychometric change, change on the Violence Risk Scale: Sexual Offender Version, and Standard Goal Attainment Scaling scores. Across the entire psychometric battery, RCZ, CSC, and RCSC were all significantly

positively correlated with SGAS scores and total change scores on the VRS:SO Dynamic scale, with correlations ranging from $r = .17, p < .05$, to $r = .28, p < .001$. SGAS and VRS:SO Dynamic change scores were strongly positively correlated, $r = .69, p < .001$. There were also several significant correlations between the individual factors of the VRS:SO and the psychometric battery, and SGAS scores. Most notably, the SGAS was correlated with all three VRS:SO factor change scores, as well as RCZ, RCSC, and CSC on the Social Inadequacy psychometric factor (F1), and RCZ and CSC on Pro-offending Attitudes (F4). Several significant correlations were also found between change scores on VRS:SO total Dynamic scale and F1 (Sexual Deviance) and psychometrics factors (RCZ, CSC, and RCSC). Overall, the correlations shown in Table 9 suggest the convergent validity of each of these separate measures of treatment outcome.

2.4.3 Relationship Between Treatment Outcome and Recidivism

The mean amount of time at risk for reoffending (from date of release to date of reoffense or follow-up) was 5.8 years for the entire sample ($N = 495$), and 4.5 years for the subsample ($N = 223$) who were rated on the SGAS and VRS:SO ($N = 218$). As reported by Allan et al. (2007), 9.9% of the sample received a reconviction for a sexual offense, 9.3% for a violent offense, and 15.7% for a general offense. Reoffending rates were slightly lower among the subsample: 7.6%, 8.1%, and 11.7% respectively. Static-99 scores were significantly correlated with sexual recidivism ($r = .27, p < .001$ for the whole sample, and $r = .22, p < .01$ for the subsample). The analyses below address the extent to which treatment outcome was related to recidivism for each of the three measures – change on the psychometric battery, change on the Violence Risk Scale: Sexual Offender Version, and Standard Goal Attainment Scaling scores.

2.4.3.1 *Psychometric change and recidivism.*

Table 10 presents the results of a correlational analysis examining the relationship between raw change scores on the psychometric tests and sexual recidivism (left columns). Change scores were calculated so that positive values indicated change in a pro-social direction, therefore the correlations with recidivism were expected to be negative (i.e., smaller pro-social change associated with recidivism). Several of the change scores were significantly correlated with recidivism. The correlations for the AI-RP and the STAXI–Control subscale were significant and in the expected direction. But correlations for the ABCS and WSFQ–Impersonal subscale were in the opposite direction to expectation (i.e., more pro-social change was related to increased risk of reoffending). This replicates the counterintuitive and somewhat troubling result reported by Hudson et al. (2002), who analyzed a subset ($N = 219$) of the present data across a shorter follow-up time. (Hudson et al. also reported a significant correlation for the WSFQ–Sadomasochistic and STAXI–Suppression change scores, but this was not replicated in our larger sample).

We next calculated correlations between the standardized change-score residuals for each test (RCZ) and sexual recidivism (i.e., partial correlations). These are listed in the right columns of Table 10, and reveal a very different pattern of association than the correlations with raw change scores (shown in the left columns). For eight variables, the partial correlations were significant and all were negative, indicating that greater pro-social change was associated with a reduction in reoffending. For five of these variables – depression (BDI), trait anxiety (STAI), trait anger and anger suppression (STAXI), and internal locus of control (ANSIE) – the

Table 10
Correlations Between Raw Psychometric Change Scores and Sexual Recidivism, and Partial Correlations (Controlling for Pre-Treatment Scores)

<i>Test</i>	<i>Change</i>		<i>Change Partial</i>	
	<i>r</i>	<i>N</i>	<i>r</i>	<i>N</i>
<i>Social Inadequacy (F1)</i>				
Social Self-Esteem Inventory ^a	.06	446	.06	445
Assertion Inventory				
Response Probability	-.14**	373	-.11*	372
State Trait Anger Expression Inv.				
Suppression	-.08	391	-.15**	390
Fear of Intimacy Scale	-.01	266	-.04	265
UCLA Loneliness	-.06	328	-.08	327
Hostility Toward Women	.05	435	-.05	434
A N-S Internal-External Control	-.04	439	-.10*	438
Beck Depression Inventory	.01	342	-.10*	340
State Trait Anxiety Inventory				
State	.01	450	-.03	449
Trait	-.03	450	-.11*	447
<i>Sexual Interests (F2)</i>				
Wilson Sex Fantasy Questionnaire				
Exploratory	.09	431	-.02	429
Intimate	.08	430	.04	429
Impersonal	.18***	430	.06	429
Sado/masochistic	.05	430	-.06	429
<i>Anger/Hostility (F3)</i>				
State Trait Anger Expression Inv.				
State	.02	397	-.05	396
Trait	-.07	399	-.16**	398
Expression	.03	398	-.04	397
Control ^a	-.10*	398	-.13**	397
<i>Pro-offending Attitudes (F4)</i>				
Abel-Becker Cognition Scale	.12**	436	-.10*	435
Rape Myth Acceptance Scale	.07	442	-.02	441
A N-S Internal-External Control	-.04	439	-.10*	438

^a For these variables, relatively higher scores were considered pro-social; hence negative correlations with recidivism were predicted. Change scores were computed as Post – Pre, such that positive values indicated pro-social change.

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

partial correlations were significant whereas the raw correlations were not. For anger control (STAXI) and assertiveness (AI-RP) both the partial and raw correlations were

significant. However, the most surprising result was obtained for the distorted cognitions scale (ABCS): Both correlations were significant but the sign of the partial correlation was reversed compared to the raw correlation. Partialling out the pre-treatment score reversed the direction of the relationship between the ABCS change score and recidivism. Moreover, the partial correlation for the WSFQ–Impersonal scale was not significant. Thus, the increased risk associated with greater pro-social change reported by Hudson et al. (2002) for two WSFQ subscales was not obtained here, when the pre-treatment scores were partialled out. The change in correlations when the pre-treatment scores were partialled exemplifies a “suppressor” effect, which occurs when a potential covariate that is unrelated to an outcome variable increases the overall model fit when added to the model (see Cohen & Cohen, 1983).

Next we calculated the relationship between recidivism and each measure of psychometric change (RCZ, CSC, and RCSC) averaged overall and for the four psychometric factors identified by Allan et al. (2007): F1 Social Inadequacy, F2 Sexual Interests, F3 Anger/Hostility, and F4 Pro-Offending Attitudes. Correlations are shown in Table 11, along with areas under the curve (AUCs) and 95% confidence intervals from Receiver Operating Characteristic (ROC) analyses. Table 11 shows that RCZ was generally related to reduced sexual recidivism, with a correlation of $r = -.11, p < .05$ for the overall average across all tests. Correlations with recidivism were also significant for F1 Social Inadequacy, F3 Anger/Hostility, and F4 Pro-offending Attitudes. Table 11 also shows that CSC, but not RCSC, was generally related to decreased sexual recidivism. The correlation for the overall average CSC across all tests was $r = -.18, p < .001$ (the corresponding correlation for overall RCSC was $r = .02, ns$). Average CSC for psychometric factors F1 Social Inadequacy, F2 Sexual Interests, and F4 Pro-offending Attitudes were also significantly negatively related to recidivism. Given that the CSC criteria do not

Table 11

Predictive Validity of Measures of Treatment Outcome: Correlations with Sexual Recidivism and AUCs for Change on a Four-Factor Psychometric Battery (RCZ, CSC and RCSC); SGAS Scores; and Change on VRS:SO Dynamic Scale and Factor Scores

<i>Measure of Treatment Outcome</i>	N	r	<i>Sexual recidivism</i>	
			AUC	AUC 95% CI
Psychometric RCZ				
F1 Social Inadequacy	458	-.09*	.59	.49 - .68
F2 Sexual Interests	431	.01	.47	.37 - .57
F3 Anger/Hostility	400	-.14**	.64**	.54 - .74
F4 Pro-offending Attitudes	458	-.09*	.58	.49 - .67
Overall RCZ	458	-.11*	.62*	.53 - .71
Psychometric CSC				
F1 Social Inadequacy	458	-.14**	.64**	.56 - .72
F2 Sexual Interests	431	-.10*	.59	.49 - .68
F3 Anger/Hostility	400	-.09	.58	.47 - .69
F4 Pro-offending Attitudes	458	-.12**	.62*	.53 - .70
Overall CSC	458	-.18***	.66***	.58 - .75
Psychometric RCSC				
F1 Social Inadequacy	458	-.03	.53	.44 - .63
F2 Sexual Interests	431	.12*	.41	.31 - .51
F3 Anger/Hostility	400	-.06	.57	.47 - .66
F4 Pro-offending Attitudes	458	.05	.45	.36 - .54
Overall RCSC	458	.05	.43	.35 - .51
VRS:SO Change				
F1 Sexual deviance	218	-.17*	.70*	.60 - .79
F2 Criminality	218	.04	.43	.29 - .57
F3 Treatment responsivity	218	.14	.60	.45 - .74
Total Dynamic Change	218	-.15*	.67*	.56 - .79
Standard Goal Attainment Scaling				
Show empathy and insight into victim issues	223	-.12	.62	.48 - .76
Accept personal responsibility for offending	223	-.11	.60	.45 - .74
Recognize cognitive distortions	223	-.10	.60	.48 - .72
Understand offense chain	223	-.10	.59	.45 - .74
Identify relapse prevention concepts	223	-.12	.61	.48 - .75
Motivation to change behavior	223	-.16*	.66*	.53 - .79
SGAS Total	223	-.16*	.66*	.52 - .80

Note. For Receiver Operating Characteristic analyses the test direction was set so that the positive value for prediction was non-recidivism, since higher scores on the measures of treatment outcome indicate better performance. Thus, AUCs in the .50 to 1.0 range represent a predictive relationship between treatment gain and reduced recidivism.

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

actually assess change on a measure, but rather whether the post-treatment scores are within a normative (non-offender) range, CSC as used here is possibly more

accurately described as a post-treatment measure of dynamic risk factors rather than a measure of change made in treatment. It is therefore unsurprising that CSC is predictive of reduced recidivism. When the level of change across treatment is also taken into account (RCSC), there was generally no relationship between change on the psychometric factors and recidivism. The exception is F2 Sexual interests, for which a correlation in the opposite direction to expected was found, $r = .12, p < .05$. A probable explanation for this counter-intuitive result relates to the positive relationship between pre-treatment and change scores discussed above, meaning that men with more deviant pre-treatment scores are able to demonstrate a greater degree of change. The correlation between F2 RSCS and recidivism is therefore likely to have been influenced by the strong relationship between pre-treatment scores on this factor and recidivism ($AUC = .72, p < .001$; as reported by Allan et al., 2007), coupled with the relative transparency of the WSFQ.

To summarize the relationship between change on the psychometric battery and sexual recidivism, significant correlations were found between partial change scores (controlling for pre-treatment scores) and decreased recidivism for several tests. This was also the case when psychometric change was averaged across four validated dynamic risk dimensions to reduce the data to the more manageable and clinically relevant factor breakdown. Finally, using reliable and clinically significant change methodology RCSC was not predictive of reduced recidivism. CSC (i.e., the extent to which post-treatment scores were indistinguishable from non-offender norms) was generally related to reduced recidivism, particularly for F1 Social Inadequacy and F4 Pro-offending Attitudes.

2.4.3.2 Violence Risk Scale: Sexual Offender Version and recidivism.

Table 11 shows correlations between VRS:SO change scores (on the Total Dynamic scale and the three VRS:SO factors) and recidivism, along with AUCs and 95% confidence intervals. Because higher change scores indicate a greater reduction in dynamic risk, negative correlations are expected if treatment gains are linked to decreased recidivism. Change scores on the Total Dynamic scale were significantly negatively correlated with sexual recidivism, $r = -.15, p < .05$, indicating that men who were rated as having made greater change in treatment were less likely to be reconvicted of a sexual offense after release. To support this finding, we also found that the sexual recidivism rate was significantly higher among those who attained the median amount of change or less on the VRS:SO Dynamic scale (10.9%), compared to those whose change scores were above the median (3.7%), $\chi^2(1) = 4.16, p < .05$. There was also a significant negative correlation between change on the Sexual Deviance factor (F1) and sexual recidivism, however correlations for the other two VRS:SO factor change scores (F2 Criminality and F3 Treatment Responsivity) were not significant.

2.4.3.3 Standard Goal Attainment Scaling and recidivism.

Correlations and AUCs examining the predictive validity of Total SGAS scores with regard to sexual recidivism are also shown in Table 11. Total SGAS scores were negatively correlated with recidivism, $r = -.16, p < .05$, indicating that those offenders who were judged as having attained the goals of treatment to a greater extent were less likely to have been reconvicted of a sexual offense at follow-up. Scores on the single SGAS goal “Motivation to change behavior” were also significantly related to recidivism, $r = -.17, p < .05$.

2.4.4 *Can Measures of Treatment Outcome Contribute to Risk Prediction Above and Beyond Static Factors?*

The above analyses showed that standardized residual change (controlling for pre-treatment scores; RCZ) and clinically significant change (CSC; post-treatment scores within the normative range) on relevant psychometric tests, attainment of in-treatment goals (SGAS), and change on an objectively-rated measure of dynamic risk (VRS:SO) were each related to reduced levels of sexual recidivism: Men who showed more improvement on these measures during treatment were less likely to reoffend sexually than men who made relatively less improvement. However, an important question is whether measures of treatment change contribute additional predictive validity for recidivism beyond static factors. If treatment change measures fail to predict recidivism after controlling for static factors, it could suggest that measures of treatment outcome are proxies for static risk, with lower risk offenders generally performing better in treatment.

Table 12 shows the results of hierarchical logistic regression analyses in which Static-99 scores were entered first, followed by each measure of treatment outcome at the second step, with sexual recidivism as the dependent variable. The columns in Table 12 showing the exponentiated coefficients (e^B) indicate the changes in risk associated with a point increase on the predictor variables, with an e^B greater than 1 meaning higher scores are related to increased recidivism, and e^B lower than 1 meaning the opposite. For example, the e^B of 1.45 for the Static-99 indicates that an increase of one point on this measure is associated with an increase of 45% in the odds of sexual recidivism.

Several of the measures of treatment outcome shown in Table 12 were significant predictors of reduced recidivism after controlling for the Static-99: RCZ for F3 Anger/Hostility ($e^B = .57$), overall CSC ($e^B = .09$) and CSC on F1 Social

Table 12

Logistic Regression Analyses of the Predictive Validity of Measures of Treatment Outcome (RCZ, CSC, SGAS, VRS:SO Change) Controlling for Static-99

<i>Measure of Treatment Outcome (correlation with Static-99)</i>	<i>Static-99</i>		<i>Treatment Outcome</i>	
	<i>B (SE)</i>	<i>e^B</i>	<i>B (SE)</i>	<i>e^B</i>
<i>Psychometric RCZ</i>				
F1 Social Inadequacy (-.05)	.38*** (.07)	1.46	-.42 (.25)	.66
F2 Sexual Interests (-.04)	.38*** (.07)	1.47	.06 (.21)	1.06
F3 Anger/Hostility (-.08)	.38*** (.08)	1.46	-.56* (.24)	.57
F4 Pro-offending Attitudes (-.03)	.38*** (.07)	1.46	-.34 (.21)	.71
Overall RCZ (-.08)	.38*** (.07)	1.46	-.60 ^a (.31)	.55
<i>Psychometric CSC</i>				
F1 Social Inadequacy (-.10*)	.37*** (.07)	1.45	-1.56* (.62)	.21
F2 Sexual Interests (-.13*)	.38*** (.08)	1.46	-.73 (.48)	.48
F3 Anger/Hostility (-.10 ^a)	.37*** (.08)	1.45	-.96 (.72)	.38
F4 Pro-offending Attitudes (-.10*)	.38*** (.07)	1.46	-1.01* (.48)	.36
Overall CSC (-.16 ^{**})	.36*** (.07)	1.44	-2.26 ^{**} (.77)	.11
<i>VRS:SO Change</i>				
F1 Sexual deviance (-.04)	.32** (.12)	1.38	-.56* (.25)	.57
F2 Criminality (.08)	.34** (.12)	1.40	.25 (.67)	1.28
F3 Treatment responsivity (.03)	.33** (.11)	1.39	-.65 (.45)	.52
Total Dynamic Change (-.01)	.32** (.11)	1.38	-.27 ^a (.14)	.76
<i>SGAS Total (-.16*)</i>				
Motivation to Change Behavior (-.11)	.34** (.12)	1.40	-.12 ^b (.06)	.89

* $p < .05$, ** $p < .01$, *** $p < .001$, ^a $p < .06$, ^b $p < .07$

Inadequacy ($e^B = .21$) and F4 Pro-offending Attitudes ($e^B = .36$), and change on the VRS:SO F1 Sexual Deviance ($e^B = .57$). Approaching but not reaching significance were overall RCZ ($e^B = .55$, $p < .06$), change scores on the VRS:SO Total Dynamic scale ($e^B = .76$, $p < .06$), SGAS Total scores ($e^B = .89$, $p < .07$), and scores on the goal “Motivation to Change Behavior” ($e^B = .59$, $p < .06$). Also shown in Table 12 are correlations between each measure of treatment outcome and the Static-99, to address the question of whether lower risk offenders performed better in treatment. Static-99 scores were not significantly correlated with RCZ or VRS:SO scores, but were related to CSC and SGAS scores. This suggests that treatment change is not a proxy for static risk, but that CSC and SGAS scores, both assessed only at post-treatment, are risk

indicators as well as measures of treatment gain. In contrast, these figures suggest that RCZ and change on the VRS:SO can be viewed as measures of change achieved in treatment that are independent of risk level.

2.5 Discussion

In this study three methods for assessing proximal treatment outcome among men who had completed a program for sexual offenders against children were evaluated in terms of their predictive validity for reoffending after release. Two of these methods involved measuring changes from pre-treatment to post-treatment on identified dynamic risk factors – one using a battery of self-report psychometric tests and the other ratings according to behavioral descriptors (VRS:SO). The third method (SGAS) also involved structured ratings, but based on post-treatment estimates of attainment of treatment goals. All measures of treatment outcome employed in this study were correlated positively with each other. The relatively strong correlation between VRS:SO change and SGAS scores ($r = .69$) was likely influenced by both scales being rated retrospectively from the same file information. Nonetheless, our results suggest the convergent validity of all of these measures of treatment outcome. Predictive validity in terms of sexual recidivism was also found for each of the three methods of measuring treatment outcome; these results are summarized below.

Change on the psychometric measures across treatment were analyzed in several different ways. Significant pro-social change was apparent from the raw difference between the pre- and post-treatment scores. However, the correlations with recidivism highlighted the problems associated with analysing raw change data, which were likely exacerbated by the transparency of some measures and the incentives for the men to demonstrate treatment gain – offenders who had more deviant pre-

treatment scores were able to show greater amounts of change (whether genuine or the result of a self-presentation bias post-treatment). In an attempt to ameliorate this problem we statistically controlled for pre-treatment scores, and averaged the standardized residual change across the entire battery and for each of the four dimensions of dynamic risk identified by Allan et al. (2007). Significant relationships were found between pro-social change and decreased sexual recidivism across the entire battery, as well as for the domains of Social Inadequacy, Anger/Hostility, and Pro-offending Attitudes.

We also computed the percentages of the sample who had achieved clinically significant change (CSC) and reliable and clinically significant change (RCSC) on the psychometric battery. Averaging across all tests and the entire sample, the probability of clinically significant change (meaning that the post-treatment score was in the normative range) on a given measure was 62.5%. This is a similar finding to that reported by Beech and Hamilton-Giachritsis (2005), who found that 66.4% of their sample (averaged across 12 treatment groups) achieved CSC on selected psychometric tests. In the present study, CSC was related to decreased sexual recidivism overall and for Allan et al.'s (2007) domains of Social Inadequacy, Sexual Interests, and Pro-offending Attitudes. A much smaller proportion of the sample (19.4%) achieved RCSC – a level of change that, as well as being clinically significant, was also reliable (i.e., beyond that attributable to test measurement variability). RCSC is perhaps more accurately labelled as a measure of gains achieved in treatment than CSC alone, which only takes into account post-treatment scores. However, RCSC was not associated with decreased sexual recidivism in the present sample.

Positive treatment outcome was associated with decreased sexual recidivism for both of the measures involving ratings based on behavioral descriptors – change

on the Violence Risk Scale: Sexual Offender Version (VRS:SO), and Standard Goal Attainment Scaling (SGAS) scores. Results for the VRS:SO suggested that positive change on the dynamic factor “Sexual Deviance” in particular was correlated with reduced recidivism. Similarly, among individual treatment goal items comprising the SGAS, “Motivation to change behavior” was most strongly predictive of recidivism. These results have implications for improving the efficacy of program content and delivery. For example, they suggest that a stronger focus on addressing the dynamic risk items that load onto the VRS:SO Sexual Deviance factor could result in further reductions in the recidivism rate of men who undergo treatment. The behavioral descriptors associated with these items include features such as a lifestyle revolved around sexual deviancy, a preference for deviant sexual stimuli, repetitive compulsive sexual behavior, and poor self-regulation, grooming or planning in the build-up to an offense (Wong et al., 2006). Treatment modalities to target these behaviors could include reconditioning of sexual arousal patterns, teaching skills for lifestyle balance and general and sexual self-regulation, and enhancement of victim empathy. Enhancing motivation to change behavior (i.e., motivation to avoid reoffending in the future) as indicated by the SGAS results could involve the incorporation of motivational interviewing techniques (Ginsburg, Mann, Rotgers, & Weekes, 2002) to encourage progression through the stages of change. Also, application of principles from a “Good Lives” approach to treatment (e.g., Ward & Stewart, 2003), specifically a focus on approach rather than avoidance treatment goals, has been shown to be related to therapist judgements of an offender’s genuine motivation to live life without further sexual offending (Mann, Webster, Schofield, & Marshall, 2004). Our results suggest this might in turn have a significant impact on reducing recidivism rates.

In the present study we found a significant relationship between positive treatment gains and reduced recidivism for three separate methods of assessing

treatment outcome (psychometric change, change on the VRS:SO, and SGAS). There are several implications of this result. In addition to providing support for the validity of the measures used to assess treatment outcome, our results also endorse the efficacy of the treatment program (see Hanson, 1997b), and most notably, provide empirical support for the premise that dynamic variables are both changeable and related to recidivism. That measurable changes on dynamic risk factors were related to reduced recidivism is especially encouraging, because relatively few studies have found significant relationships between treatment outcome and recidivism (e.g., Barbaree, 2005; Quinsey et al., 1998; exceptions include Hudson et al., 2002, who found this general trend however with inconsistencies between psychometric tests; and Olver et al., 2007). It is also notable that treatment gain was related to reduced recidivism in the present sample of relatively low risk sex offenders (average Static-99 score in the moderate-low range), because previous studies have found such correlations only among the higher risk offenders in their sample (Marques et al., 2005; Olver et al., 2007).

In terms of the comparative predictive validity of the measures of treatment outcome, Table VI shows that RCZ, CSC, VRS:SO, and SGAS were all predictive of sexual recidivism, with AUCs for overall scores ranging from .62 for RCZ to .67 for VRS:SO change. Overall CSC (AUC = .66) was the only measure with an overall score that predicted significant incremental variance beyond that predicted by the Static-99 (Table 12). However, given that CSC is a measure of post-treatment deviancy rather than treatment gain per se (as noted above), this finding is perhaps best viewed as contributing to the growing evidence that dynamic factors can make an independent contribution to risk prediction beyond static factors (e.g., Allan et al., 2007; Beech et al., 2002; Beggs & Grace, 2008 [Study 1, this volume]; Hanson & Harris, 2000; Olver et al., 2007).

Regarding the other measures of treatment outcome, the predictive ability of the SGAS (AUC = .66) is especially impressive given the relative simplicity and efficiency of the measure. To illustrate, rating each file on the SGAS required approximately 15-25 minutes (including file review; a therapist who had been working with the case could likely apply the ratings more quickly still). Change on the VRS:SO Dynamic scale had comparable predictive validity (AUC = .67), however completing the ratings was a much more complex and time-consuming process (up to approximately 90 minutes per file). Similarly, to administer, score, and analyze an extensive battery of psychometrics at two points in time in order to calculate the standardized residual change (AUC = .62) is considerably less efficient. Table 12 indicates that the SGAS also approached significance in predicting recidivism even after controlling for the variance predicted by the Static-99. In practical terms, the odds ratio of $e^B = .89$ means that an increase of one point on the Total SGAS scale (range -12 to +12) is associated with an 11% decrease in the odds of sexual recidivism. Therefore, the results of this study suggest that the SGAS can be recommended for a quick and valid assessment of proximal treatment outcome that is predictive of reduced recidivism.

However, the VRS:SO holds other advantages over the SGAS. In addition to its function as a measure of treatment outcome, a pre-treatment assessment using the VRS:SO can be informative about important treatment targets or criminogenic needs for a particular offender. Motivational ratings (an important responsivity issue) are also assessed for each identified criminogenic need. Additionally, the VRS:SO was designed with the predominant function of risk assessment, and scores on the measure have been validated as being significantly predictive of sexual recidivism with AUC values for pre- and post-treatment total scores in the range of .71 to .79 (Beggs & Grace, 2008; Olver et al., 2007). Thus, a wealth of clinical and risk information can be

gained using the VRS:SO aside from treatment change. The same is also true of the battery of psychometric tests, which can provide clinically useful information as to the individual's emotional functioning, interpersonal competence, and sexual attitudes and beliefs, and significantly added to the predictive validity of the Static-99 (Allan et al., 2007).

In summary, this study provides evidence that specific gains made in treatment are significantly correlated with a reduction in recidivism risk. Three separate methods of assessing such gains were compared: Change across treatment on a battery of psychometric tests (analyzed in several different ways); change on the dynamic scale of the VRS:SO; and SGAS scores. Convergent validity was found between these methods, and each was related to decreased sexual recidivism. SGAS is recommended as a brief screen of treatment outcome among sex offenders. Both the VRS:SO and the psychometric battery provide an abundance of clinically useful information and have been shown elsewhere to be more strongly related to recidivism than treatment outcome alone was in this study. The VRS:SO has the additional advantage of encompassing all of these features within a single tool, which has been independently validated.

Study 3. Psychopathy, Treatment Goal Attainment, and Sex Offender Recidivism

3.1 Abstract

We examined the relationship between treatment outcome, psychopathy, and recidivism among 223 completers of Kia Marama, a prison-based program for child sex offenders. Treatment outcome was assessed using a modified version of Hogue's (1994) Standard Goal Attainment Scaling for Sex Offenders rating system, and psychopathy was measured with the Psychopathy Checklist - Revised (PCL-R; Hare, 1991). Higher PCL-R scores were associated with overall increased risk of reoffending (sexual, violent, and serious), whereas positive treatment outcome was correlated with decreased sexual recidivism. No evidence was found for an interaction between psychopathy and treatment outcome for any type of recidivism. Separate analyses performed for incest and extrafamilial offenders found that the relationship between treatment outcome and recidivism risk varied depending on the type of sex offender. For incest offenders, positive treatment outcome was associated with an increased risk of violent, but not sexual, recidivism, whereas for extrafamilial offenders positive treatment outcome was associated with a reduced risk of both sexual and violent recidivism.

There is a current controversy about whether the Psychopathy Checklist - Revised (PCL-R; Hare, 1991, 2003) can be used to identify a group of sexual offenders for whom treatment is not only ineffective, but actually might increase the likelihood of reoffending. Although earlier studies suggested that treated psychopaths had higher recidivism rates than non-treated psychopaths (Ogloff, Wong, & Greenwood, 1990; Rice, Harris, & Cormier, 1992), a report by Seto and Barbaree (1999) attracted considerable attention because it appeared to show that sex offenders who scored high on the PCL-R were more likely to reoffend if they showed good rather than poor treatment behavior.

Seto and Barbaree (1999) explored the relationship between psychopathy, treatment behavior, and recidivism in a sample of 224 men released from a prison-based treatment program in Canada. Clinical case notes and treatment reports were rated on eight items measuring both in-group behavior (e.g., level of disruptiveness, appropriateness of interactions) and treatment change (change in victim empathy, understanding of offense cycle, quality of relapse prevention plan). An overall measure of treatment behavior was obtained by averaging the individual items. During the follow-up period ($M = 32$ months), 7.6% of the sample committed a new serious offense (a violent or sexual offense). Multiple regression analyses found that both treatment behavior and psychopathy were significantly related to serious recidivism, although the relationship for treatment behavior was in the unexpected direction (with good treatment behavior being associated with increased recidivism). Because the interaction term approached significance, Seto and Barbaree conducted a further analysis in which the offenders were divided into four groups based on median-splits of PCL-R and treatment behavior scores. They found that the group of

offenders who had relatively high PCL-R scores (above the sample median of 15) and showed good treatment behavior were the most likely to commit a new serious offense.

The impact of Seto and Barbaree's (1999) article was substantial. It has been widely cited – a Web of Science search in July 2006 obtained 69 hits - more than any other article with “sexual offenders” as a keyword ($N = 49$) and in the top 1.3% of all articles published in psychological journals in 1999 ($N = 2,296$). The questions raised by Seto and Barbaree's report are fundamentally important for the treatment and management of sex offenders: Is treatment outcome related to reoffending, and is there a subgroup of offenders with psychopathic traits for whom treatment might be harmful? If the latter is true, it could lead to the practice of excluding offenders who obtain high PCL-R scores from treatment programs.

However, recent results suggest that Seto and Barbaree's (1999) conclusions may not be warranted. Barbaree (2005) conducted an extended follow-up of the same sample of offenders as Seto and Barbaree. In addition to a longer follow-up time ($M = 5.2$ years), Barbaree obtained a more comprehensive set of recidivism records from a Canadian national Police database. Using the more comprehensive data, when the offenders were again divided into the same four groups by performing median-splits on treatment behavior and PCL-R scores, there was no evidence of an interaction between psychopathy and treatment behavior. Men with relatively high PCL-R scores (> 15) were more likely to reoffend, but there was no relationship between treatment behavior and recidivism. These results were confirmed in additional analyses based on Cox regression and fixed follow-up periods. Barbaree concluded that the more complete data failed to replicate the major results reported in 1999, and that their data provided no evidence that participation in treatment could increase the recidivism risk

of offenders with relatively high PCL-R scores. The discrepancy between the 1999 and 2005 results was explained in terms of the more comprehensive offense records from the Police database. Further support for Barbaree's (2005) conclusions was more recently provided in a study by Langton et al. (2006), who applied a modified version of the Seto and Barbaree (1999) treatment behavior measure to a larger and updated dataset subsuming the samples of both Seto and Barbaree (1999), and Barbaree (2005). Langton et al.'s averaged "Response to Treatment" score was found to be unrelated to serious or sexual recidivism.

Looman, Abracen, Serin, and Marquis (2005) conducted a further investigation of psychopathy, treatment outcome, and recidivism. They studied 154 consecutive admissions to the Regional Treatment Centre Sex Offender Treatment Program (RTCSOTP) in Ontario, but recidivism data were only available for 102 offenders. Compared to Seto and Barbaree (1999), Looman et al.'s sample had overall higher PCL-R scores; the mean was 22.54, near the cutoff of 25 for diagnosis of psychopathy recommended by Quinsey, Harris, Rice and Cormier (1998). Looman et al. also used two different measures related to treatment performance. The first was a measure of treatment behavior in which clinical case notes and post-treatment reports were rated on victim awareness, quality of offense cycle, and relapse prevention plans; the second was a yes/no "risk rating" which indicated whether the offenders' risk was judged by a therapist to have been reduced by participation in treatment. The risk rating was determined through a structured process and used multiple sources of information, including an actuarial measure (Violence Risk Appraisal Guide; VRAG; Quinsey et al., 1998) and reports from nurses, corrections officers and recreation staff.

To examine the possible interaction between treatment outcome and psychopathy on recidivism, Looman et al. (2005) conducted median-split analyses

similar to Seto and Barbaree (1999). Using treatment behavior as the measure of treatment outcome, they found that the high PCL-R, good treatment behavior group had a significantly higher rate of serious recidivism compared to the two low PCL-R groups. Looman et al. noted that this was an apparent replication of Seto and Barbaree's result. However, results were different when groups were dichotomized on the basis of the "risk rating" – the high PCL-R, no risk reduction group offended at a higher rate than the two low PCL-R groups. Moreover, there was no significant difference between the high PCL-R, risk reduction group and the two low PCL-R groups, leading Looman et al. to suggest that there might be a group of individuals who score high on psychopathy and yet respond well to treatment. They speculated that such a group might not perform well on measures of treatment behavior because of resistance (Mahoney, 1991). According to this view, highly antisocial individuals may show increased levels of disruptive behavior if treatment is causing pro-social change. If so, traditional measures of treatment behavior (which typically involve compliance with therapists' instructions) may not be appropriate for individuals with high PCL-R scores. Whether or not this hypothesis is tenable, the results of the Looman et al. study, taken together with those of Barbaree (2005) and Langton et al. (2006), present a conflicting picture. The PCL-R appears to be a valid predictor of recidivism risk for sex offenders, but whether measures of treatment outcome correspond to reductions in risk, and whether the PCL-R can be used to identify offenders who will not respond well to treatment, remains unclear.

The present study represents a further attempt to explore the relationship between psychopathy, treatment outcome, and recidivism. We obtained records for 223 men who completed the Kia Marama program, a prison-based treatment program for sexual offenders against children in New Zealand (Hudson, Marshall, Ward,

Johnston, & Jones, 1995; Hudson, Wales, & Ward, 1998). There were three major differences between our study and previous research (Barbaree, 2005; Langton et al., 2006; Looman et al., 2005; Seto & Barbaree, 1999). Our sample consisted entirely of child molesters (i.e., rapists were excluded), and was divided approximately equally between incest and extrafamilial offenders. In this way, we planned to compare results for different subtypes of child molesters. Second, our sample included a greater proportion of low and moderate risk offenders, and consequently PCL-R scores were lower overall than those reported by the aforementioned previous studies. Thus, our study provided an opportunity to test the generality of the relationship between PCL-R scores, treatment outcome and recidivism in a population with a somewhat lower risk profile than previous research. Finally, our measure of treatment outcome was derived from the Standard Goal Attainment Scaling (SGAS) methodology developed by Hogue (1994) for sex offenders. The original SGAS methodology is a 12-factor rating system which measures the extent to which treatment goals have been realized. It has been used in previous studies to assess treatment impact and motivation to change at pre-treatment, post-treatment and community follow-up for low-moderate and high risk offenders (Barrett, Wilson & Long, 2003; Stirpe, Wilson & Long, 2001). Compared to the measures of treatment behavior used by Seto and Barbaree (1999), Barbaree (2005), Looman et al. (2005), and Langton et al. (2006), the SGAS methodology provides a potentially more detailed and systematic assessment of treatment change, as it includes a greater number of clinical change items with a wider rating scale, while not including items pertaining to compliance or conduct in group.

3.3 Method

3.3.1 Participants

The sample consisted of 223 males who completed the treatment program at the Kia Marama Special Treatment Unit between 1993 and 2000. Kia Marama is a 32-week, prison-based group treatment program for men who have sexually offended against children, and is located in a separate unit in a medium security prison near Christchurch, New Zealand. The program is based on cognitive-behavioral principles, with an emphasis on relapse prevention, and is organized as a therapeutic community. Groups of eight men attend two and a half hour therapy sessions three times per week, covering the following modules: Norm Building, Understanding your Offending, Arousal Reconditioning, Victim Impact and Empathy, Mood Management, Relationship Skills, and Relapse Prevention. All participants gave written consent for their file information to be used for research and evaluation purposes when entering the program. For a more thorough description of the Kia Marama program, see Hudson et al. (1995).

3.3.2 Measures

3.3.2.1 Psychopathy.

The Psychopathy Checklist - Revised (PCL-R; Hare, 1991) was administered to all men as part of the standard post-treatment assessment at Kia Marama. The PCL-R consists of 20 items rated on a three-point scale ranging from 0 to 2, and has demonstrated reliability and validity. According to Hare (1991), psychopathy is most usefully viewed as a dimensional construct, with the total score reflecting the extent that the individual matches the description of a prototypical psychopath. For situations

in which a categorical conceptualization is required, Hare recommended a cut-off total score of 30 or higher as appropriate for diagnosis of a psychopathic personality. However, subsequent cross-cultural research (Hare, Clark, Grann & Thornton, 2000) has found that a cut-off score of 25 may be more valid among Western cultures other than North American, and has been recommended by Quinsey et al. (1998).

3.3.2.2 Treatment outcome.

Treatment outcome in the short-term was assessed following completion of the program using a modified version of Hogue's (1994) standardized Goal Attainment Scaling (SGAS) for sex offenders. This measure consisted of six goals of treatment for which participants were rated on their level of attainment. Attainment of each goal was rated on a five-point scale ranging from -2 to +2 (with a score of zero representing a minimally acceptable level of goal attainment), resulting in a Total SGAS score ranging from -12 to +12. The goals were: 1) Show empathy and insight into victim issues; 2) Accept personal responsibility for offending; 3) Recognise cognitive distortions; 4) Understand offense chain; 5) Identify relapse prevention concepts; and 6) Motivation to change behavior. SGAS scores were rated by two data coders from post-treatment reports on file written by the group therapist. The coders were blind as to recidivism outcomes while conducting the ratings, and completed these independently of each other. A randomly selected ten percent of the sample ($N = 23$) were rated by both coders, and the reliability was acceptably high (Spearman's ρ for SGAS Total score = .82, $p < .01$) and comparable to previous studies that have used the SGAS (Barrett et al., 2003; Stirpe et al., 2001).

3.3.2.3 Recidivism.

Reconviction data as of January 2001 were obtained from the official criminal history database maintained by the NZ Department of Corrections, giving follow-up times ranging from twelve to 90 months following treatment. Data for sexual, violent, and general offenses were recorded separately.

3.4 Results

3.4.1 Demographic and Offense Information

The age of men in the sample ranged from 18 to 74 years, with an average of 41.06 ($SD = 11.90$). Educational attainment, recorded as part of a demographic questionnaire completed on program entry, was broken down into five levels. Nine percent of the sample had a primary education only; 36% had completed up to two years, 34% three years, and 10% between four and five years of secondary school; with the remaining 11% completing some tertiary study. Over half (56.5%) of the sample were incest offenders whose victims were exclusively family members, whereas 43.5% were extrafamilial offenders with at least one unrelated victim.

The mean amount of time at risk for reoffending (from date of release to date of reoffense or end of follow-up time) was 54 months ($M = 4.5$ years), ranging from 36 days to 90 months. Of the 223 men, 18.9% had received a new conviction of any kind at the end of the follow-up period. The percentages of men who received convictions for sexual, violent, and general offenses were 7.6%, 8.1%, and 11.7% respectively.

3.4.2 Psychopathy and Treatment Outcome

PCL-R scores for the sample ranged from 0 to 33. The mean score for the sample was 8.23 ($SD = 7.28$), and the median was 7. Using the cutoff score of 25 discussed by Hare et al (2000) and used by Looman et al. (2005), 4.5% of the sample could be diagnosed as psychopaths (or 2.2% using a cutoff score of 30). Because of the relatively low PCL-R score distribution in this sample, it is important to clarify that the term “psychopathy” is used in the remainder of this chapter to refer to the dimensional construct of traits, as opposed to a clinical cut-off.

Overall, the sample performed satisfactorily in attaining treatment goals. Almost two thirds (63 %) received a positive Total SGAS score, indicating at least a minimally acceptable level of goal attainment. The average Total score was 0.59 ($SD = 4.14$). Table 13 shows the average scores for each of the six SGAS goals, which ranged from $M = -0.21$, $SD = 1.00$ for showing empathy and insight into victim issues, to $M = 0.37$, $SD = 0.81$ for understanding offense chain. Table 13 also presents these analyses separately for incest and extrafamilial offenders; these will be discussed in a later section.

Table 13.
Means and Standard Deviations of Age, PCL-R Scores and SGAS Scores

	Whole Sample (N=223)		Incest (N=126)		Extrafamilial (N=97)	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Age	41.06	11.90	41.01	10.99	41.12	13.03
PCL-R	8.23	7.28	7.29*	6.44	9.45*	8.12
SGAS -						
<i>Show empathy & insight into victim issues</i>	-.21	1.00	-.05**	1.04	-.41**	.91
<i>Accept personal responsibility</i>	-.12	.87	-.01*	.85	-.27*	.88
<i>Recognise cognitive distortions</i>	-.06	.76	-.01	.76	-.13	.76
<i>Understand offense chain</i>	.37	.81	.38	.77	.35	.87
<i>Identify relapse prevention concepts</i>	.35	1.02	.38	1.01	.29	1.04
<i>Motivation to change behavior</i>	.27	.99	.35	.98	.18	1.01
SGAS TOTAL	.59	4.14	1.04	4.13	.01	4.11

* $p < .05$ ** $p < .01$ *** $p < .001$ significant differences between incest and extrafamilial offenders

Correlational analyses between the demographic variables, PCL-R scores, and SGAS scores are shown in Table 14. There was a significant relationship between PCL-R scores and age, $r = -.23, p < .01$, such that those scoring higher on the PCL-R tended to be younger. However, the relationship between the PCL-R and education level was not significant, $r = -.10, ns$. There was no relationship between age or education level on Total SGAS scores. However, an examination of the individual SGAS goals indicated that older offenders were less likely to have accepted responsibility for their offending, $r = -.17, p < .05$, while those with higher levels of education were more likely to have gained an understanding of their offense chain, $r = .19, p < .01$.

Psychopathic traits as measured by the PCL-R were negatively related to Total SGAS scores, $r = -.34, p < .001$, as well to all six individual treatment goals. After program completion, offenders with relatively high PCL-R scores within the sample were less likely to show empathy and insight into victim issues, $r = -.32, p < .001$, accept personal responsibility for their offending, $r = -.25, p < .001$, recognize cognitive distortions, $r = -.22, p < .01$, understand their offense chain, $r = -.21, p < .01$, identify relapse prevention concepts, $r = -.20, p < .01$, and be motivated to change their behavior, $r = -.34, p < .001$. Thus, higher PCL-R scores were associated with negative treatment outcomes in general.

3.4.3 Recidivism

Correlational analyses were carried out to examine the relationships between recidivism and demographic variables, psychopathy, and treatment outcome (Table 15). Correlations between education level and recidivism were not significant. Age was not significantly related to sexual recidivism, $r = -.10, ns$, but was negatively

Table 14.
Correlations – Demographic Variables, PCL-R, and SGAS

	Age	Education	PCL-R	Empathy	Responsibility	Cognitive distortions	SGAS Offense chain	Relapse prevention	Motivation	SGAS TOTAL
Age										
Education	-.10									
PCL-R	-.23**	-.10								
SGAS - Empathy	-.08	.05	-.32***							
Responsibility	-.17*	.00	-.25***	.56***						
Cognitive distortions	-.09	.01	-.22**	.34***	.52***					
Offense chain	-.03	.19**	-.21**	.41***	.41***	.32***				
Relapse prevention	-.07	.13	-.20**	.49***	.48***	.40***	.53***			
Motivation	-.03	.07	-.34***	.59***	.55***	.39***	.60***	.64***		
SGAS TOTAL	-.10	.10	-.34***	.77***	.77***	.63***	.71***	.80***	.84***	

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

Table 15.
Correlations – Demographic Variables, PCL-R, SGAS, and Recidivism

	Whole Sample (N=223)			Incest (N=126)			Extrafamilial (N=97)		
	<i>Sexual</i>	<i>Violent</i>	<i>General</i>	<i>Sexual</i>	<i>Violent</i>	<i>General</i>	<i>Sexual</i>	<i>Violent</i>	<i>General</i>
Age	-.10	-.26***	-.35***	-.14	-.29**	-.45***	-.07	-.24*	-.24*
Education	-.07	-.06	-.02	-.03	-.02	.09	-.13	-.10	-.13
PCL-R	.24***	.24***	.33***	.28**	.24**	.29**	.20	.28**	.39***
SGAS -									
<i>Empathy</i>	-.12	.06	-.04	-.01	.17	.04	-.20	-.17	-.18
<i>Responsibility</i>	-.11	.05	.05	-.09	.21*	.20*	-.09	-.21*	-.17
<i>Cognitive distortions</i>	-.10	.00	.03	-.10	.11	.16	-.09	-.18	-.17
<i>Offense chain</i>	-.10	-.07	-.08	.08	.02	-.03	-.24*	-.20*	-.14
<i>Relapse prevention</i>	-.12	.04	-.04	-.03	.19*	.09	-.20	-.20	-.23*
<i>Motivation</i>	-.16*	-.03	-.07	-.04	.09	.04	-.24*	-.24*	-.23*
SGAS TOTAL	-.16*	-.04	-.09	-.04	.18*	.11	-.24*	-.27**	-.25*

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

correlated with general recidivism, $r = -.35, p < .001$, and violent recidivism, $r = -.26, p < .001$, with younger offenders being more likely to have been convicted of a violent or general reoffense at follow-up. PCL-R scores were found to predict sexual, $r = .24, p < .001$, violent, $r = .24, p < .001$, and general, $r = .33, p < .001$, recidivism, consistent with prior studies that have demonstrated that psychopathy is a valid predictor of reoffense risk (Hemphill, Hare, & Wong, 1998).

Short-term treatment outcome, as measured by Total SGAS scores, was negatively related to sexual recidivism, $r = -.16, p < .05$, but not to violent, $r = .01, ns$, or general recidivism, $r = -.04, ns$. Additionally, the SGAS goal of motivation to change behavior was found to be negatively correlated with sexual recidivism, $r = -.16, p < .05$. Thus, offenders who had overall better treatment outcomes, and who were more motivated to change their behavior, were less likely to have been reconvicted of a sexual offense at follow-up.

3.4.3.1 *Survival analyses.*

We conducted Cox regression analyses to examine the effects of psychopathy (PCL-R scores), treatment outcome as measured by the Total SGAS score, and their interaction on recidivism, while taking into account time at risk. To be comparable with Barbaree (2005) and Looman et al. (2005), serious recidivism (defined as a reconviction for either a sexual or violent offense) was used as the outcome variable. Separate analyses were also conducted for sexual and violent recidivism, given that we found a significant correlation between treatment outcome and sexual recidivism, but not violent.

As can be seen in Table 16, results revealed a significant PCL-R effect for sexual, $\beta = .10, p < .01$, and violent recidivism, $\beta = .13, p < .001$, as well as for

Table 16.
Cox Regression Results for tests of PCL-R, SGAS, and Interaction on Recidivism

	Whole Sample (N=223)		Incest (N=126)		Extrafamilial (N=97)	
	β	$Exp(\beta)$	β	$Exp(\beta)$	β	$Exp(\beta)$
<i>Serious Recidivism</i>						
Total SGAS	.02	1.03	.16*	1.17	-.14	.87
PCL-R	.10***	1.11	.13***	1.14	.09	1.09
Interaction	.00	1.00	-.01	1.00	.00	1.00
<i>Sexual Recidivism</i>						
Total SGAS	-.10	.99	.08	1.08	-.16	.85
PCL-R	.10**	1.10	.12*	1.13	.10*	1.10
Interaction	.00	1.00	-.02	.98	.01	1.01
<i>Violent Recidivism</i>						
Total SGAS	.12	1.13	.21*	1.24	-.19	.83
PCL-R	.13***	1.13	.13**	1.14	.13	1.14
Interaction	.00	1.00	.00	1.00	.01	1.01

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

serious recidivism, $\beta = .10$, $p < .001$. However, Total SGAS scores did not have a significant effect on rates of any kind of recidivism, and the change in χ^2 when the interaction term was added was also not significant, $\chi^2 (df = 1) = .105$, *ns* for serious and violent recidivism; $\chi^2 (df = 1) = .227$, *ns* for sexual. Thus, the Cox regression analyses confirmed that PCL-R scores were positively associated with recidivism, but there was no evidence of an interaction between treatment outcome and recidivism.

3.4.3.2 Median-split analyses.

To examine further the relationship between treatment outcome, PCL-R scores, and recidivism we conducted a median-split analysis. Each case was assigned to one of four groups according to whether they scored above or below the median on the PCL-R ($Md = 7$) and Total SGAS ($Md = 1$). Table 17 shows the resulting sample size and recidivism rates (sexual, violent, and serious) for each group. Again, it is important to note that the low PCL-R score median means that the “High” and “Low” labels refer only to relative levels of psychopathic traits within the present sample.

Table 17.
Group Sizes and Recidivism Rates Following Median-Splits of PCL-R (High/Low) and Total SGAS (Good/Poor) Scores

Group	N	% Recidivism		
		<i>Serious</i>	<i>Sexual</i>	<i>Violent</i>
<i>Entire Sample (N = 223)</i>				
High-Good	40	20.0	7.5	15.0
High-Poor	73	19.2	13.7	12.3
Low-Good	65	6.2	1.5	4.6
Low-Poor	45	6.7	6.7	0.0
<i>Incest Offenders (N = 126)</i>				
High-Good	27	25.9	7.4	22.2
High-Poor	32	12.5	6.2	9.4
Low-Good	38	10.5	2.6	7.9
Low-Poor	29	3.4	3.4	0.0
<i>Extrafamilial Offenders (N = 97)</i>				
High-Good	17	5.9	5.9	0.0
High-Poor	31	29.0	25.8	16.1
Low-Good	32	0.0	0.0	0.0
Low-Poor	17	17.6	11.8	5.9

Kaplan-Meier survival analyses were used to compare the recidivism rates of the four groups while taking into account differences in time at risk. Overall, there were significant between-group differences (Generalized Wilcoxon, $df = 3$) for sexual, $\chi^2 = 12.16, p < .01$, violent, $\chi^2 = 9.29, p < .05$, and serious recidivism, $\chi^2 = 12.85, p < .01$. For serious recidivism (Figure 4), a PCL-R effect is evident, with high PCL-R groups having faster reoffense rates than low PCL-R groups. This was also the case for violent recidivism. Pairwise comparisons ($df = 1$) revealed significant differences between each High PCL-R group and each Low PCL-R group for serious recidivism, ranging from $\chi^2 = 4.50, p < .05$ between High-Good and Low-Poor, to $\chi^2 = 7.77, p < .01$ between High-Poor and Low-Good. For violent recidivism, pairwise differences were significant between Low-Poor and both High PCL-R groups, $\chi^2 = 8.06, p < .01$ for High-Good, and $\chi^2 = 6.02, p < .05$ for High-Poor.

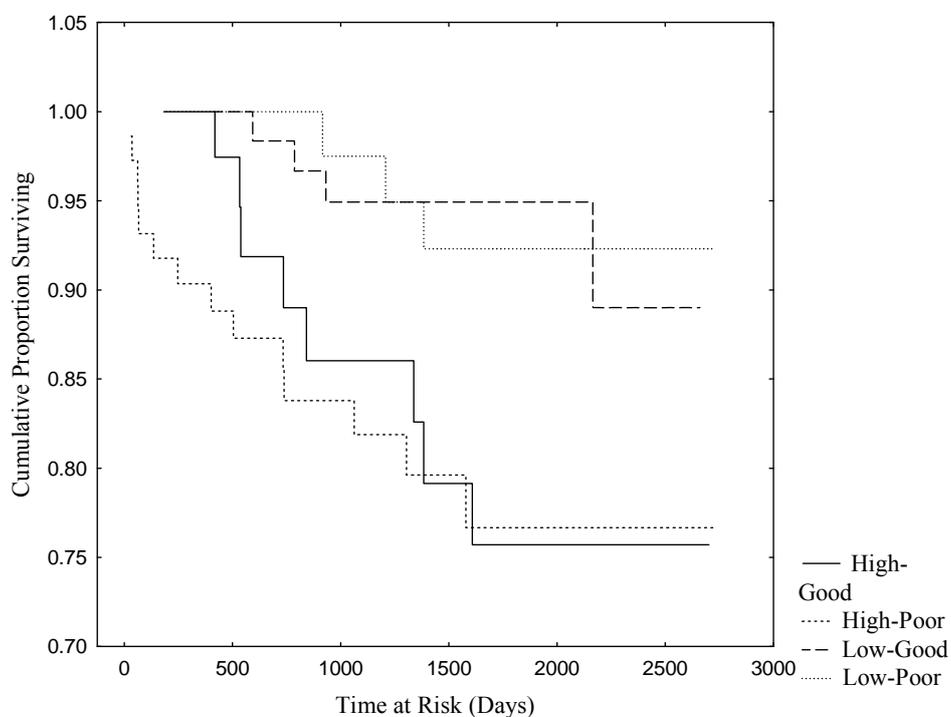


Figure 4. Kaplan-Meier survival plot for serious recidivism for PCL-R and Total SGAS median-split groups.

For sexual recidivism (Figure 5), the group with the fastest reoffense rate was High-Poor, which was significantly faster than the Low-Good group, $\chi^2 = 9.56, p < .01$. No other pairwise comparisons for sexual recidivism were significant. The lack of significant differences in sexual recidivism rates between the High-Good group and either of the Low PCL-R groups may suggest that a positive treatment response in this sex offender program was able to ameliorate the expected effects of a higher than average PCL-R score on sexual recidivism.

Overall, these results are similar to those reported by Barbaree (2005) and Looman et al. (2005), and do not replicate those found by Seto and Barbaree (1999). We did not find any link between positive treatment outcomes and increased recidivism, regardless of PCL-R scores. The strongest predictor of recidivism was psychopathy, particularly for serious and violent recidivism.

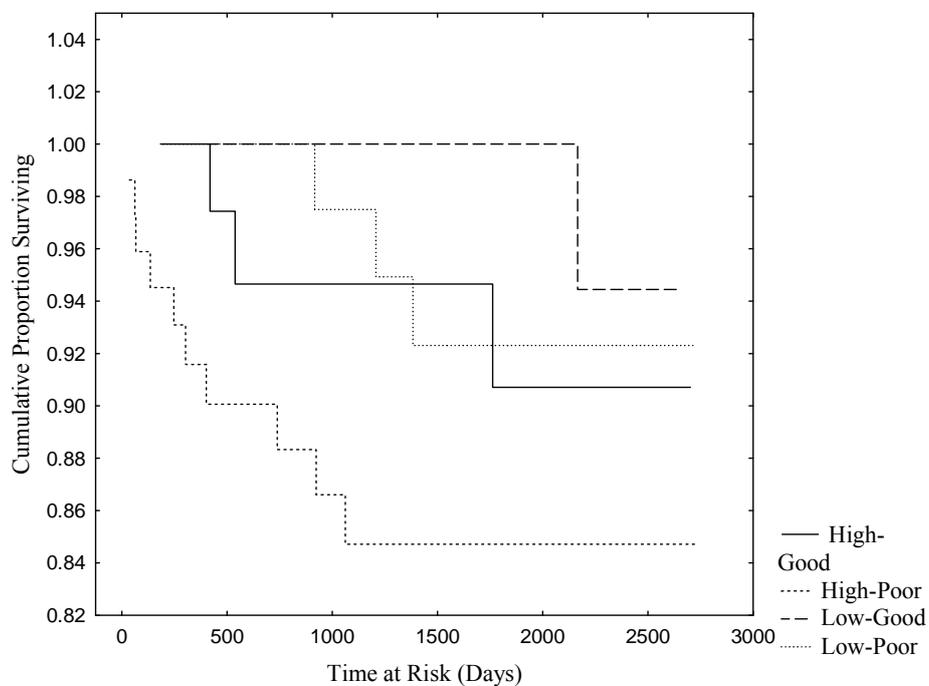


Figure 5. Kaplan-Meier survival plot for sexual recidivism for PCL-R and Total SGAS median-split groups.

3.4.4 Incest and Extrafamilial Offenders Compared

One explanation for the mixed results reported here and in previous research (cf. Seto & Barbaree, 1999; Barbaree, 2005; and Looman et al, 2005) regarding the effects of psychopathy and treatment outcome on reoffending could be related to within-group differences in sex offender samples. As Table 13 shows, there were significant differences between incest and extrafamilial offenders in our sample. Although the two groups did not differ in terms of age, extrafamilial offenders had higher levels of education than incest offenders, and also scored higher on the PCL-R. Treatment outcomes were more positive among incest offenders, who performed better than extrafamilial offenders on the SGAS goals of showing empathy and insight into victim issues, and accepting personal responsibility for offending. The difference between the two groups on Total SGAS scores approached significance, with incest offenders performing overall better ($M = 1.04$, $SD = 4.13$) than extrafamilial offenders ($M = 0.01$, $SD = 4.11$), $t(221) = 1.86$, $p = .06$.

Extrafamilial offenders also had higher rates of sexual reconviction at follow up (11.3%) than incest offenders (4.8%), a difference which approached significance despite the reduced sample sizes, $\chi^2 (df = 1) = 3.37$, $p = .07$. As shown in Table 15, the relationships between PCL-R and SGAS scores and recidivism also differed between these two subgroups. PCL-R scores were positively related to all types of recidivism among incest offenders, but among extrafamilial offenders, the correlation with sexual recidivism was not significant. Among incest offenders, Total SGAS scores were unrelated to sexual and general recidivism, but were positively correlated with violent recidivism, such that incest offenders who performed well at attaining the treatment goals were actually more likely to reoffend violently after release. However,

among extrafamilial offenders, Total SGAS scores were negatively related to both sexual and violent recidivism, as well as general recidivism.

Thus, the extrafamilial offenders in our sample tended to be better educated, have more psychopathic traits, respond less positively to treatment, and be more likely to reoffend sexually after release, compared to incest offenders. In addition, correlational data suggest that the relationships between psychopathy, treatment outcome and recidivism may have differed between these groups. To explore this possibility, we replicated our survival and median-split analyses separately for subsamples of incest and extrafamilial offenders.

3.4.4.1 Survival analyses.

Cox regression results are presented in Table 16. Similar to results for the entire sample, no significant interaction was found between Total SGAS and PCL-R scores on serious, sexual, or violent recidivism, for either incest or extrafamilial offenders. For incest offenders, the coefficient for SGAS scores was significant for serious and violent recidivism, but not sexual recidivism. However, the direction of the effect was opposite to expectation, with higher SGAS scores (indicating a more positive treatment response) associated with increased serious and violent recidivism. PCL-R effects were significant for all three recidivism types for incest offenders, as was true for the sample as a whole. Among extrafamilial offenders, there was a significant PCL-R effect for sexual recidivism only, although the effect approached significance for serious recidivism ($\beta = .09, p = .06$). A treatment effect in the expected direction (higher SGAS scores being linked to lower recidivism) approached significance for serious ($\beta = -.14, p = .07$) and sexual recidivism ($\beta = -.16, p = .06$).

3.4.4.2 Median-split analyses.

Repeating the procedure of creating four groups based on median-splits of PCL-R and Total SGAS scores on our sub-samples of incest ($N = 126$) and extrafamilial ($N = 97$) offenders yielded the results shown in Table 17. Patterns of reoffending across the four groups differed dramatically between incest and extrafamilial offenders, most notably between the Good SGAS and Poor SGAS groups. Among incest offenders, the group with the highest rate of all recidivism types was High-Good. Although there were no significant differences between the groups for sexual recidivism among incest offenders (for which recidivism rates were relatively low across all groups), Kaplan-Meier pairwise comparisons (Generalized Wilcoxon, $df = 1$) were significant between High-Good and Low-Poor groups for serious, $\chi^2 = 5.18, p < .05$, and violent recidivism, $\chi^2 = 6.47, p < .05$. These results are consistent with the Cox regression results reported above and suggest that positive treatment outcome as measured by the Total SGAS score is associated with increased serious and violent recidivism among incest offenders.

By contrast, for extrafamilial offenders recidivism rates were greatest in the High-Poor group, and those with poor treatment outcomes had higher rates of recidivism than those performing well for both Low and High PCL-R groups. For all three types of recidivism, Kaplan-Meier pairwise differences ($df = 1$) were significant between High-Poor and Low-Good ($\chi^2 = 11.29, p < .01$ for serious; $\chi^2 = 9.86, p < .01$ for sexual; and $\chi^2 = 5.05, p < .05$ for violent recidivism). The High-Poor group also had a higher rate of serious recidivism, $\chi^2 = 4.12, p < .05$, and sexual recidivism, $\chi^2 = 3.62, p = .06$, than Low-Poor, indicating that PCL-R scores significantly predicted recidivism among those extrafamilial offenders not performing well in treatment. Correspondingly, the lack of significant differences in recidivism between High-Good

and Low-Good ($\chi^2 = 2.00$, *ns*) for serious and sexual recidivism (no comparison could be made for violent recidivism as 0% of both Good SGAS extrafamilial groups reoffended violently) may suggest that PCL-R scores are not predictive among extrafamilial offenders who do attain the treatment goals. Additionally, among extrafamilial offenders the difference between Low-Good and Low-Poor was significant for serious recidivism, $\chi^2 = 4.72$, $p < .05$ (with Low-Poor being higher). The difference between these two groups approached significance for sexual recidivism, $\chi^2 = 3.23$, $p = .07$, as did the difference between High-Poor and High-Good for serious recidivism, $\chi^2 = 3.34$, $p = .07$.

Overall, the Cox regression and median-split analyses suggest that the relationship between treatment outcome, as measured by the SGAS, and recidivism is different for incest and extrafamilial offenders. For incest offenders, better treatment outcome was associated with increased violent recidivism risk, whereas for extrafamilial offenders the relationship was in the expected direction: Higher SGAS scores were associated with lower risk of all types of recidivism. However, for both groups the PCL-R was a valid predictor of reoffending.

3.5 Discussion

We investigated the relationship between psychopathy, treatment outcome, and recidivism in a sample of child molesters who completed a prison-based treatment program. Psychopathy, as measured by the PCL-R, was a valid predictor of sexual, violent, and general recidivism. Positive treatment outcome, as measured by the Standard Goal Attainment Scaling methodology (SGAS; Hogue, 1994), was negatively related to sexual recidivism, but not violent or general recidivism. However, we found no evidence of an interaction between PCL-R scores and

treatment outcome on any type of recidivism. Unlike results of Seto and Barbaree (1999), offenders with PCL-R scores higher than the sample median were not more likely to reoffend if they had good treatment outcome.

Our results are consistent with Barbaree's (2005) extended follow up of the sample studied by Seto and Barbaree (1999). For that sample, as well as the present data, PCL-R scores were a strong predictor of reoffense risk. Our results also support the findings of Langton et al. (2006), and Looman et al.'s (2005) median-split analyses using their risk reduction measure of treatment outcome. For sexual recidivism, our median-split analyses mirrored theirs – among offenders scoring relatively high on psychopathy, those with poor treatment outcomes were significantly more likely to reoffend than those who scored low on psychopathy; while those who had positive treatment outcomes did not differ from low PCL-R scorers in terms of recidivism. This is an important finding, because it suggests that treatment may be beneficial for some high-risk offenders, although it should be noted that PCL-R scores for our sample were lower compared with those studied by Langton et al. (2006), Looman et al. (2005), and Barbaree (2005). Overall, the present results add to a growing body of research suggesting that PCL-R scores do not necessarily determine offenders' capability to benefit from treatment in terms of reduced recidivism.

Looman et al. (2005) did report one set of results that appeared to replicate Seto and Barbaree (1999). Using a measure of treatment behavior, they found that offenders scoring high on psychopathy and good on treatment behavior had increased recidivism. They suggested that an explanation for their mixed findings could relate to the inclusion of items related to compliance in the treatment behavior measure. Specifically, resistance to change might be evidenced by some offenders for whom treatment is actually effective, producing lower scores on measures that are focused

on compliance and cooperation with authority. The present study is arguably consistent with Looman et al.'s explanation, given that the SGAS goals primarily pertain to clinical change, not in-group behavior. According to Hanson, Cox, and Woszczyzna (1991), appropriate measures of improvement in treatment should be based on reductions in factors that are associated with recidivism. By that definition, the measures of treatment outcome used in past studies may not have been valid assessments of treatment success. Barbaree's (2005) and Looman et al.'s ratings of treatment behavior had no relationship with recidivism, whereas good ratings on Seto and Barbaree's measure were significantly correlated with increased recidivism. By contrast, Total SGAS scores in the present study were related to reductions in sexual recidivism. This result supports the validity of Hogue's (1994) SGAS methodology for assessing treatment outcome among sex offenders (Barrett et al., 2003; Stirpe et al., 2001).

Another aspect of Seto and Barbaree's (1999), and Looman et al.'s (2005) results (as well as Barbaree, 2005, who failed to find a treatment effect), was their use of "serious recidivism" as the dependent variable, defined as a reconviction for a sexual or violent offense. In the present study, there were important differences between results for sexual and violent recidivism. Most notably, treatment effects were in opposing directions, with good treatment outcome associated with reduced sexual recidivism but increased violent recidivism. Given that none of these prior studies examined sexual and violent recidivism separately, it seems possible that their use of serious recidivism as the dependent variable may have masked an actual effect of treatment behavior on sexual recidivism. Looman et al. cited Quinsey et al.'s (1998) argument that combining sexual and violent convictions provides a more realistic measure of sexual recidivism because many sexual offenses are recorded as

lesser violent charges due to the practice of plea bargaining. However, the extent to which this takes place in New Zealand is difficult to assess, and the importance of analysing sexual recidivism separately from violent has previously been emphasised by Marques, Day, Nelson, and West (1994) and Hanson and Bussière (1998). Our results suggest that sexual and violent offenses should be examined separately, even if they are also combined as serious recidivism. Langton et al. (2006) examined both sexual recidivism and the more inclusive serious recidivism category, and found that ratings of treatment response were not significantly predictive of either outcome.

Looman et al. (2005) noted that subgroups of psychopaths may be differentially responsive to treatment. We found some evidence to support this view. For our sub-sample of incest offenders, high attainment of treatment goals was associated with increased risk for serious and violent recidivism, with the highest rates for those who also scored relatively high on psychopathy (however this result was not obtained for sexual recidivism, or for extrafamilial offenders). There is no clear explanation for this reverse treatment effect. Firestone et al. (1999) reported that incest offenders who reoffended violently had higher scores on the Michigan Alcohol Screening Test (MAST) than non-reoffenders. Speculatively, perhaps incest offenders with high SGAS scores (in particular on the goals “accept personal responsibility for offending,” and “show insight and empathy into victim issues” for which they tended to score higher than extrafamilial offenders) may experience more guilt for their past offending. Having lost (in many cases) the support of the family in which they offended, these offenders may turn to maladaptive coping strategies that are linked to violence, such as substance abuse (Boles & Miotto, 2003). Such an effect may be limited to incest offenders, for whom sexual offending in the first place may have

represented maladaptive coping, as it often occurs during times of high life stress or dissatisfaction (Hartley, 2001).

Despite the apparent increase in violent recidivism, comparatively few incest offenders committed a sexual reoffense (6 individuals out of 126, or 4.8%). This is encouraging, given that the ultimate goal of the Kia Marama program is to reduce sexual reoffending; violent behavior is not specifically targeted. Possibly, mere participation in the program is effective at reducing sexual recidivism among incest offenders, and this objective is achieved regardless of the men's performance in attaining specific treatment goals. The lack of relationship between SGAS scores and sexual recidivism among incest offenders may therefore be due to a floor effect. Another possibility is that once released, incest offenders are less likely to reoffend sexually because their contact with former victims is greatly limited. According to the risk principle (Andrews & Bonta, 2003), treatment programs (and especially intensive treatment) should be targeted towards higher risk offenders, and may be unnecessary, ineffective, or even harmful for lower risk offenders. Perhaps the paradoxical relationship in the present study for incest offenders between positive treatment outcome and increased violent recidivism may have been the result of providing intensive treatment for a low-risk sample. To examine this possibility, violent and sexual recidivism rates for treated and untreated incest offenders would need to be compared. By contrast, the robust negative relationship between Total SGAS scores and recidivism among extrafamilial offenders suggests that the SGAS is a valid measure of treatment outcome, and may be useful in contributing to post-treatment risk decisions for this population.

Overall, the present study contributes to recent research suggesting that PCL-R scores are not determinative of offenders' potential to benefit from treatment in terms

of reduced recidivism (Barbaree, 2005; Looman et al., 2005). It also suggests that the relationship between psychopathic traits, treatment outcome and recidivism may depend on the offender subgroup, as well as recidivism type.

Study 4. Psychopathy, Intelligence and Recidivism in Child Molesters: Evidence of
an Interaction Effect

4.1 Abstract

We studied the relationships between psychopathy, intelligence, and offending in a sample of treated child molesters ($N = 216$). Regression analyses showed that psychopathy (as measured by the Psychopathy Checklist – Revised; Hare, 1991) was strongly related to both offense history and recidivism during follow-up. Intelligence (assessed using 4-subtest short forms of the Wechsler Adult Intelligence Scale – Revised and 3rd versions; Reynolds, Willson, & Clark, 1983; Wechsler, 1999) was not related to offending. However, there was a significant interaction between intelligence and psychopathy on recidivism: Offenders with relatively low intelligence and high psychopathy scores were more than four times as likely to have received a sexual reconviction as other offenders. Results are discussed in terms of implications for risk assessment.

Psychopathy is a personality construct that is robustly correlated with criminal behavior (Hare, Clark, Grann & Thornton, 2000). As measured by the Psychopathy Checklist – Revised (PCL-R; Hare, 1991), psychopathic personality has been linked to a host of negative outcomes, including an increased amount of offending, particularly violent offending (Hare, 1991), increased institutional misconducts (Guy, Edens, Anthony, & Douglas, 2005), and higher rates of recidivism after release (Hemphill, Hare & Wong, 1998). Personality traits associated with psychopathy include superficiality, grandiosity, impulsivity, emotional shallowness, and a lack of remorse or empathy (Hare, 1991).

Intelligence is also thought to be related to criminality, a view that has received mixed support in the literature dating back several decades. In a review of the early research, Hirschi and Hindelang (1977) found that low intelligence was associated with delinquent behavior as measured by both official records and self reports. In addition, they found that intelligence was a better predictor of delinquency than social class or race. Subsequently, in a meta-analysis Gendreau, Little and Goggin (1996) reported a significant mean effect size of $r = .07$ between low intelligence and recidivism among adult offenders across 32 studies ($N = 21,369$).

Guay, Ouimet and Proulx (2005) discussed two possible hypotheses for the relationship between intelligence and criminal behavior. The first, endorsed by Hirschi and Hindelang (1977), is that intelligence is indirectly linked to offending through its effect on mediating factors such as school and job performance, adaptation, and opportunities for pro-social success. Although it is apparent that intelligence would be related to school performance, Fergusson and Horwood (1995) reported data from a longitudinal birth cohort study suggesting that there is a strong

correlation between cognitive factors and externalizing behaviors throughout development, and that this relationship can explain the later correlation between academic achievement and delinquency. The second hypothesis refers to a direct relationship, in which people with lower intelligence are more prone to crime because of their weakness in relevant cognitive abilities such as anticipating the consequences of their actions and recognizing suffering in others (Guay et al., 2005). But note that individuals with such weaknesses might also be described as having a lack of remorse or empathy, or behaving impulsively – traits which are associated with psychopathy. This, together with the evidence that both variables are related to criminal behavior, raises the question of what effect psychopathy and intelligence may have in combination. In particular, researchers (e.g., Heilbrun, 1979, 1982; Walsh, Swogger & Kosson, 2004) have been concerned with whether intelligence might moderate the effect of psychopathy on criminal behavior, specifically whether psychopathic personality might be a stronger predictor of offending for individuals with relatively low intelligence compared to those with relatively high intelligence.

In an early study with incarcerated offenders, Heilbrun (1979) found evidence for such an interaction between psychopathy and intelligence on violent and impulsive offending. Heilbrun divided a sample of incarcerated offenders into four groups based on median splits on measures of psychopathy and intelligence, and found that the group with higher psychopathic scores and lower IQ scores had more previous violent and impulsive offenses than those with high psychopathy and high IQ scores and both low psychopathy groups. Heilbrun (1982) proposed that a combination of poor impulse control, low empathy, and negative socialization was implicated in the interaction between psychopathy and intelligence.

A problematic aspect of Heilbrun's studies is their reliance on various self-report methodologies to assess psychopathy. Thus it is important to re-examine the relationship between psychopathy, intelligence, and offending using a modern, well-validated instrument such as the PCL-R. Recently, Walsh et al. (2004) attempted to replicate Heilbrun's (1979) results with a sample of jail inmates serving short sentences (less than one year) using the PCL-R. Regression analyses found main effects for psychopathy and intelligence in postdicting the number of past violent offenses of the sample, but the interaction was not significant. Comparisons of extreme groups (formed by grouping upper and lower thirds of IQ and PCL-R scores) showed that among European Americans (but not African Americans), low IQ psychopathic offenders had significantly more charges for violent offenses (including sexual violence) than any other group, although Walsh et al. noted that this appeared to reflect an additive effect of psychopathy and intelligence on violence, rather than interactive. They also suggested that Heilbrun's results could be understood in terms of additive effects of psychopathy and intelligence (see also Holland, Beckett, & Levi, 1981).

It is possible that factors that have been demonstrated to be predictive of increased recidivism in general (e.g., intelligence, and psychopathy) may not necessarily be predictive among sexual offenders. Sexual offenders have long been recognized as a special subtype of offenders, because the etiology of their criminal behavior may involve particular factors (e.g., sexual deviance) that do not apply to offenders in general (Hanson & Bussière, 1998; Hanson & Morton-Bourgon, 2005; but cf. Miethe, Olson & Mitchell, 2006). In recent years, risk assessment for sex offenders has emerged as a major area of research, and various actuarial instruments have been developed which have good validity for predicting recidivism, such as the

Static-99 (Hanson & Thornton, 1999; see Beech, Fisher & Thornton, 2003, for review). Typically, these instruments include a number of offense-history variables, known as static risk factors, which have been shown to be related to sexual reoffending (e.g., number of prior sexual offenses, male victims, stranger victims). The static factors are summed to yield a total score which serves as the basis of risk prediction.

Barbaree, Seto, Langton and Peacock (2001) compared the utility of five actuarial instruments including the Static-99, and the PCL-R for predicting recidivism in sex offenders. Recidivism was defined as a conviction for a new sexual offense, serious offense, or any offense during the follow-up period ($M = 4.5$ years), and instruments were assessed in terms of areas under the Receiver Operating Characteristic curve (AUC). They found that two actuarial measures - the Static-99 and Sex Offender Risk Appraisal Guide (SORAG; Quinsey, Harris, Rice, & Cormier, 1998) – were the best overall predictors of recidivism, with average AUC values of .77, .71, and .68 for any, serious, and sexual recidivism, respectively. However, the PCL-R significantly predicted serious and any recidivism (AUC's = .63 and .68), and approached significance for sexual recidivism (AUC = .61, $p < .07$). The predictive validity of the PCL-R is not surprising because it was positively correlated with the actuarial instruments; indeed, the SORAG includes the PCL-R as an important component.

The goal of the present study was to investigate whether intelligence moderated the relationship between psychopathic personality, as measured by the PCL-R, and recidivism in child molesters. Participants were men who had received treatment at the Kia Marama program for adult sexual offenders against children and were then released to the community. We also conducted a parallel set of analyses to

examine the relationship between intelligence, actuarial risk (as measured by the Static-99), and recidivism. The rationale for this was to determine whether any possible interactive effect was specific to the interpersonal and behavioral features of psychopathy, or also applied to actuarial risk.

4.3 Method

4.3.1 Participants

The sample consisted of all 216 males who completed the Kia Marama program at Rolleston Prison, Christchurch, New Zealand, between 1993 and 2000 for which both a PCL-R score and an IQ score was available. Kia Marama is a 32 week prison-based treatment program based on cognitive-behavioral and relapse prevention principles. Participants attend three hour group treatment sessions three times per week, and reside in the adjacent therapeutic community unit. A more thorough description of this program can be found in Hudson, Marshall, Ward, Johnston and Jones (1995). All participants gave written consent for their file information to be used for research and evaluation purposes when entering the program.

4.3.2 Measures

4.3.2.1 Psychopathy.

Subjects were scored on the PCL-R as part of a post-treatment assessment. The PCL-R was designed to measure psychopathy in offender populations, and has demonstrated reliability and validity (e.g., Hare, 1991). Hare reported that the inter-rater reliability of total scores among a pooled sample of male prison inmate was .91. The PCL-R consists of 20 items rated on a three-point scale ranging from 0 to 2, resulting in a maximum total score of 40. Hare (1991) argued that psychopathy is

most usefully viewed as a dimensional construct, with an offender's total score reflecting the extent that they match the description of a prototypical psychopath. For situations in which a categorical conceptualization is required, a cut-off score of 25 has support as being appropriate for diagnosis of a psychopathic personality (Hare et al., 2000; Rice, Harris & Cormier, 1992). A dimensional interpretation of psychopathy was used in this study; therefore the term "psychopathy" refers to the level of psychopathic traits present according to PCL-R scores, rather than a diagnosis.

4.3.2.2 Intelligence.

Intelligence testing is conducted routinely at Kia Marama as part of the pre-treatment assessment battery, for the purpose of ascertaining ability to cope with treatment content. The entry guidelines of the program require an IQ of 70 or above. IQ was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), and prior to the WASI's development, a four-subtest short version (Picture Completion, Block Design, Information, and Arithmetic) of the WAIS-R (Reynolds et al., 1983). Both of these tests have demonstrated good reliability and validity as screening tools for estimating full-scale intelligence (Wechsler, 1999; Reynolds et al., 1983).

4.3.2.3 Actuarial risk.

The Static-99 (Hanson & Thornton, 1999) is an instrument designed to assess risk of recidivism in sex offenders. It consists of 10 items primarily relating to offense history, each of which is rated on a 0-1 or 0-3 scale. The total maximum score is 12, with higher scores reflecting more risk. The Static 99 has consistently been shown to

have shown good predictive accuracy for sexual recidivism (Barbaree et al., 2001). Static-99 scores were previously rated from file information for all participants. As a reliability check, two raters had independently scored a subset of 10 cases; agreement for these cases was 100%.

4.3.2.4 *Offending.*

The number of previous convictions for sexual offenses and other non-trivial offenses (not including convictions for which the current sentence was received) were recorded for each participant during the pre-treatment assessment phase. The information was obtained during an interview and confirmed against official records. Information on sexual, violent, and general reconvictions as of January 2001 (between one and seven years after completion of treatment, with an average follow-up time of 5.0 years) was obtained from the official criminal history database maintained by the NZ Department of Corrections.

4.4 *Results*

4.4.1 *Demographic and Offense Information*

The men were aged between 18 and 74 years, with an average of 41.1 ($SD = 12.0$). In terms of ethnicity, 75.9% ($n = 164$) were of New Zealand European descent, 21.8% ($n = 47$) identified as New Zealand Maori, and 2.3% ($n = 5$) other. Educational attainment was correlated with IQ, $r = .50$, $p < .001$, and ranged from primary only (9%; $n = 19$) to some tertiary (10%; $n = 22$), with the majority in between having completed some secondary school: 36% ($n = 78$) up to two years; 34% ($n = 73$) three years; and 11% ($n = 24$) between four and five years. Over half (57%; $n = 123$) of the

sample were incest offenders whose victims were exclusively family members, whereas 43% ($n = 93$) were extrafamilial offenders with at least one unrelated victim.

Table 18 shows the average number of prior sexual and other convictions for the whole sample, and for incest and extrafamilial offenders separately. The average number of prior sexual convictions of the sample was 1.4 ($SD = 3.5$), ranging from 0 (66% of the sample; $n = 142$) to a maximum of 28. Forty-eight percent ($n = 104$) had no prior convictions for other offenses. The maximum was 220 prior convictions, and the average was 8.6 ($SD = 24.8$). Regarding recidivism, 19.4% ($n = 42$) of the 216 men had received a new conviction of any kind by the end of the follow-up period after being released from prison. The percentages of men who received convictions for new sexual, violent, and general offenses were 7.9% ($n = 17$), 8.3% ($n = 18$), and 12.0% ($n = 26$) respectively. Of those convicted of a new sexual offense, the average time between release and offending was approximately 2 years (23.9 months), and ranged from 36 days to 5.9 years.

Table 18.
Means and Standard Deviations for Test Scores, Prior Convictions, and Recidivism

	Whole Sample ($N = 216$)		Incest ($n = 124$)		Extrafamilial ($n = 92$)	
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>
PCL-R	8.2	7.3	7.2*	6.4	9.5*	8.1
IQ	98.9	14.8	96.8*	14.1	101.7*	15.4
Static-99	2.2	2.0	1.3***	1.5	3.5***	2.0
<i>Prior convictions</i>						
Sexual	1.4	3.5	0.7**	2.1	2.5**	4.7
Other	8.6	24.8	6.1	13.8	12.0	34.2
<i>Recidivism</i>		% (n)		% (n)		% (n)
Sexual		7.9 (17)		4.8 (6)		12.0 (11)
Violent		8.3 (18)		9.7 (12)		6.5 (6)
General		12.0 (26)		12.9 (16)		10.9 (10)

* $p < .05$ ** $p < .01$ *** $p < .001$ significant differences between incest and extrafamilial offenders

4.4.2 *Psychopathy, Intelligence, and Actuarial Risk*

Average PCL-R, IQ, and Static-99 scores for the sample and for incest and extrafamilial offenders separately are also shown in Table 18. The average PCL-R score of the sample was 8.2 ($SD = 7.3$). Scores ranged from 0 to 33, and the median was 6.5. Using a cut-off score of 25, 4.6% ($n = 10$) of the sample could be diagnosed as psychopaths. This is lower than most prison populations (Hare, Hart, & Harpur, 1991), and is likely due to the sample consisting entirely of child molesters, who typically have lower PCL-R scores than rapists and non-sexual offenders (e.g., Porter et al, 2000). Because of the relatively low distribution of PCL-R scores in this sample, it is important to emphasize that the term “psychopathy” is used here to refer to the dimensional construct of traits, rather than the clinical cut-off. There was a significant relationship between PCL-R scores and age, $r = -.22, p < .01$, with higher PCL-R scorers tending to be younger.

The IQ scores of the sample showed a wide range from 67 to 161 (three offenders who scored under the entry criteria cut-off of 70 were granted admission based on judgements of their ability to cope with the program requirements), with a mean of 98.9 ($SD = 14.8$), and a median of 99.5. Older offenders tended to have higher IQ scores, $r = .23, p < .01$. There was a negative correlation between IQ and PCL-R scores, $r = -.20, p < .01$.

Static-99 scores ranged between 0 and 9, with a mean score of 2.2 ($SD = 2.0$), and a median of 2. The sample was therefore in the Moderate-Low risk category overall (Hanson & Thornton, 1999). Static-99 scores were negatively correlated with age, $r = -.21, p < .01$, but were unrelated to IQ scores, $r = -.03, ns$. There was a positive correlation between Static-99 and PCL-R scores, $r = .40, p < .001$.

4.4.3 Offending

Correlations between PCL-R, IQ, and Static-99 scores and the offense variables (prior sexual and other convictions, and sexual, violent, and general recidivisions) are shown in Table 19. PCL-R scores were positively related to previous sexual offenses, $r = .18, p < .01$, and previous other offenses, $r = .40, p < .001$, as well as all three types of recidivism: sexual, $r = .25, p < .001$, violent, $r = .25, p < .001$, and general, $r = .34, p < .001$. These results are consistent with prior studies that have demonstrated that psychopathy is a robust predictor of recidivism (Hemphill et al., 1998). IQ scores were unrelated to prior sexual and nonsexual offending, and to general recidivism. However, correlations with sexual and violent recidivism were significant or approached significance, such that those offenders with relatively low IQ were more likely to be reconvicted for a violent offense, $r = -.14, p < .05$, and the correlation for sexual recidivism approached significance, $r = -.13, p < .07$. As expected, Static-99 scores correlated strongly with prior sexual convictions, $r = .63, p < .001$, and also prior other convictions, $r = .30, p < .001$. Static-99 scores were also significantly related to sexual recidivism, $r = .23, p < .01$, and general recidivism, $r = .16, p < .05$, and approached significance for violent recidivism, $r = .13, p < .07$.

Table 19.
Correlations Between PCL-R, IQ, and Static-99 Scores and Offense Variables

	Prior Convictions		Recidivism			PCL-R	IQ
	Sexual	Other	Sexual	Violent	General		
PCL-R	.18**	.40***	.25***	.25***	.34***		-.20**
IQ	.05	-.09	-.13	-.14*	-.11	-.20**	
Static-99	.63***	.30***	.23**	.13	.16*	.40***	-.03

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

4.4.3.1 *Is there an interaction effect between psychopathy and intelligence on offending?*

We conducted hierarchical regressions in which prior offending behavior was first predicted by PCL-R and IQ scores, and then the interaction term was added to the model. For number of prior sexual convictions, there was a significant effect of PCL-R score, $\beta = 0.20$, $p < .01$, but neither IQ nor the interaction term reached significance, $\beta = 0.09$ and $\beta = 0.002$, respectively, both *ns*. Similar results were obtained for number of prior non-sexual convictions: The PCL-R coefficient was significant, $\beta = 0.37$, $p < .001$, whereas the IQ and interaction coefficients were not, $\beta = -0.01$ and $\beta = -0.04$, both *ns*. Thus, higher PCL-R scores were associated with increased levels of prior offending, both sexual and non-sexual, but IQ was unrelated to offense history.

Next we used recidivism during follow up as the outcome variable in a similar set of regression analyses. For sexual recidivism, the PCL-R was a significant predictor, $\beta = 0.19$, $p < .01$, but IQ was not significant, $\beta = -0.09$, *ns*. The PCL-R x IQ interaction contributed significant additional variance, $\beta = -0.16$, $R^2_{inc} = .02$, $p < .05$. Similar results were obtained with other recidivism variables. For violent recidivism, the main effect of PCL-R and the PCL-R x IQ interaction were significant, $\beta = 0.19$, $p < .05$ and $\beta = -0.16$, $R^2_{inc} = .02$, $p < .05$, respectively, but the effect of IQ was not, $\beta = -0.10$, *ns*. The PCL-R and PCL-R x IQ interaction were significantly related to general recidivism, $\beta = 0.26$, $p < .001$ and $\beta = -0.16$, $R^2_{inc} = .03$, $p < .05$, respectively, but IQ was not, $\beta = -0.05$, *ns*. Thus, for each type of recidivism, higher PCL-R scores were associated with increased risk of reoffending and there was a significant interaction between PCL-R and IQ.

Table 20.

Group Sizes and Recidivism Rates of the Total Sample and Four Groups Following Median-Splits of IQ and PCL-R Scores

	<i>n</i>	Mean Scores		% Recidivism		
		<i>PCL-R</i>	<i>IQ</i>	<i>Sexual</i>	<i>Violent</i>	<i>General</i>
Total Sample	216	8.2	98.9	7.9	8.3	12.0
<i>Group</i>						
High PCL-R-High IQ	47	12.5	110.4	4.3*	8.5	10.6
High PCL-R-Low IQ	61	14.6	87.0	18.0	18.0	24.6
Low PCL-R-High IQ	61	2.5	110.9	3.3*	3.3*	9.8*
Low PCL-R-Low IQ	47	3.0	87.2	4.3*	2.1*	0.0*

* Significantly lower rate of recidivism than the High PCL-R-Low IQ group according to Kaplan-Meier pairwise comparisons

To examine the interaction between psychopathy and intelligence on recidivism more closely, we dichotomized both PCL-R and IQ scores by performing median splits (median PCL-R score = 6.5; median IQ score = 99.5). Offenders were then assigned to one of four groups: High PCL-R-High IQ, High PCL-R-Low IQ, Low PCL-R-High IQ, and Low PCL-R-Low IQ (note that the means for the two “High PCL-R” groups, 12.5 and 14.6, are well below traditional cut-offs for a psychopathic diagnosis; the label therefore only refers to relatively high levels of psychopathic traits within this sample). Table 20 shows the number of cases, mean PCL-R and IQ scores, and the recidivism rates for sexual, violent, and general reoffending for each group. Kaplan-Meier survival analyses were used to compare the recidivism rates of the four groups while taking into account possible differences in time at risk. The percentage of cases that reoffended differed significantly across groups for sexual, $\chi^2 = 14.85, p < .01$, violent, $\chi^2 = 14.53, p < .01$, and general recidivism, $\chi^2 = 18.53, p < .001$ (all *df*'s = 3). For each type of recidivism, reconviction rates for the High PCL-R groups were greater than for the corresponding Low PCL-R groups, and the High PCL-R-Low IQ group had the highest reconviction rates of all. Kaplan-Meier pairwise comparisons (*df* = 1) confirmed that this group

was significantly more likely to have been reconvicted of a sexual offense than the other three groups: $\chi^2 = 5.01, p < .05$ for High PCL-R-High IQ; $\chi^2 = 8.24, p < .01$ for Low PCL-R-High IQ; and $\chi^2 = 5.67, p < .05$ for Low PCL-R-High IQ. The High PCL-R-Low IQ group also had a higher rate of violent recidivism than either Low PCL-R group: $\chi^2 = 7.88, p < .01$ for Low PCL-R-High IQ; and $\chi^2 = 7.48, p < .01$ for Low PCL-R-Low IQ. For general recidivism, the High PCL-R-Low IQ group differed significantly from the Low PCL-R-High IQ group, $\chi^2 = 5.94, p = .05$ and the Low IQ-Low PCL-R group, $\chi^2 = 14.20, p = .001$, and approached significance with regard to the High PCL-R-High IQ group, $\chi^2 = 3.83, p < .06$.

In summary, a significant interaction effect was found for predicting recidivism during follow-up, with a combination of high PCL-R and low IQ scores being linked to significantly greater sexual, violent, and general recidivism. By contrast, there was no interaction between PCL-R and IQ scores when postdicting prior sexual or nonsexual offenses.

4.4.3.2 *Is the interaction specific to psychopathy?*

To examine the possibility that the interaction effect found between psychopathy and intelligence on recidivism is not specific to psychopathy but reflects a more general interaction between intelligence and overall risk level, we carried out the regression analyses and median split comparisons again, but using the Static-99 instead of the PCL-R. For sexual recidivism, the Static-99 was a significant predictor, $\beta = 0.24, p < .001$, but neither IQ nor the interaction term reached significance, $\beta = -0.11, ns$, and $\beta = -0.11, R^2_{inc} = .01, ns$, respectively. Results were similar for general recidivism, with Static-99 having a significant main effect, $\beta = 0.16, p < .05$, but neither IQ, $\beta = -0.10, ns$, nor the interaction, $\beta = -0.11, ns, R^2_{inc} = .01, ns$, were

significant predictors. For violent recidivism, the Static-99 and IQ both approached significance as main effects, $\beta = 0.13, p < .06$ and $\beta = -0.12, p < .07$ respectively, and the interaction term was significant, $\beta = -0.16, p < .05, R^2_{inc} = .03, p < .05$.

Table 21 shows the percentages of sexual, violent, and general recidivism for four groups formed by dichotomizing the Static-99 and IQ scores using median splits (median Static-99 score = 2; median IQ score = 99.5); group sizes and mean scores are also given. Kaplan-Meier analyses across the four groups ($df = 3$) indicated no significant overall group differences: $\chi^2 = 7.32, ns$ for sexual; $\chi^2 = 5.02, ns$ for violent; and $\chi^2 = 7.27, ns$ for general recidivism. These results indicate that the above findings with regard to the PCL-R are not reflective of a general interaction between intelligence and risk on recidivism.

Table 21.
Group Sizes and Recidivism Rates of the Total Sample and Four Groups Following Median-Splits of IQ and Static-99 Scores

	<i>n</i>	Mean Scores		% Recidivism		
		<i>Static-99</i>	<i>IQ</i>	<i>Sexual</i>	<i>Violent</i>	<i>General</i>
Total Sample	216	8.2	98.9	7.9	8.3	12.0
<i>Group</i>						
High Static-99-High IQ	56	3.6	111.5	5.4	3.6	10.7
High Static-99-Low IQ	65	3.6	86.4	13.8	13.8	20.0
Low Static-99-High IQ	52	0.4	109.9	1.9	7.7	9.6
Low Static-99-Low IQ	43	0.5	88.1	9.3	7.0	14.0

4.4.3.3 Predictive validity.

The interaction between psychopathy and intelligence suggests that the predictive validity of the PCL-R for recidivism might depend on the offender's IQ. Table 22 shows the results of a comparative analysis in which the predictive validity of the PCL-R and Static-99 for each type of recidivism, as measured by the area under the receiver operating characteristic curve (AUC), was computed separately for

relatively high and low IQ offenders and the full sample. For the PCL-R, predictive validities for the full sample were statistically significant for each type of recidivism, and increased for the low IQ group, with AUC values of .76, .82 and .86 for sexual, violent, and general recidivism, respectively. By contrast, the Static-99 did not significantly predict violent or general recidivism for the full sample, and the AUC value for sexual recidivism was about the same for the low IQ group as for the sample as a whole. Thus the PCL-R showed greater predictive validity for recidivism than the Static-99, particularly for low IQ offenders.

Table 22.

Areas Under the Receiver Operating Characteristic Curve (AUC) for the PCL-R and Static-99 Predicting Sexual, Violent, and General Recidivism

	<i>n</i>	AUC for Recidivism		
		<i>Sexual</i>	<i>Violent</i>	<i>General</i>
<i>Full Sample</i>				
PCL-R	216	.73**	.71**	.71***
Static-99	216	.69**	.58	.62
<i>Low IQ</i>				
PCL-R	108	.76**	.82***	.86***
Static-99	108	.68*	.70*	.71*
<i>High IQ</i>				
PCL-R	108	.60	.53	.53
Static-99	108	.73	.40	.53

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

4.4.3.4 Incest and extrafamilial offender comparisons.

Table 18 illustrates several important differences between the groups of incest (N = 124) and extrafamilial (N = 92) offenders in our sample. Compared to incest offenders, extrafamilial offenders had significantly greater PCL-R scores overall, $t(168.45) = 2.22, p < .05$, and Static-99 scores, $t(161.87) = 8.74, p < .001$, and had a greater average number of previous convictions for sexual offenses, $t(117.65) = 3.58, p < .01$. Consistent with this increased risk, extrafamilial offenders also had a higher rate of sexual recidivism (12.0%) than incest offenders (4.8%), a difference that

approached significance despite the reduced sample sizes, $\chi^2 (df = 1) = 3.69, p < .06$. IQ scores were also significantly greater for extrafamilial offenders, $t(186.54) = 2.39, p < .05$. For these reasons, it is worth exploring the interaction between PCL-R scores and IQ scores on sexual recidivism for these two groups separately.

Table 23.

Group Sizes and Recidivism Rates of the Total Sample and Four Groups Following Median-Splits of IQ and PCL-R Scores

Group	Incest ($n = 124$)				Extrafamilial ($n = 92$)			
	<i>n</i>	<i>Mean Scores</i> <i>PCL-R</i>	<i>Mean Scores</i> <i>IQ</i>	<i>%</i> <i>Recidivism</i>	<i>n</i>	<i>Mean Scores</i> <i>PCL-R</i>	<i>Mean Scores</i> <i>IQ</i>	<i>%</i> <i>Recidivism</i>
High PCL-R-High IQ	29	10.8	107.0	0.0	21	14.6	113.3	9.5
High PCL-R-Low IQ	37	12.1	85.3	10.8	24	17.5	87.7	29.2
Low PCL-R-High IQ	31	1.9	108.8	3.2	27	3.5	113.9	7.4*
Low PCL-R-Low IQ	27	2.3	87.7	3.7	20	2.6	89.7	0.0*

*Significantly lower rate of recidivism than the High PCL-R-Low IQ group according to Kaplan-Meier pairwise comparisons

Table 23 shows the samples sizes, mean PCL-R and IQ scores, and percentage of sexual reconvictions for the groups resulting from median splits conducted separately for incest offenders and extrafamilial offenders. The median PCL-R and IQ scores were 6.0 and 96.0 respectively for incest offenders, and 7.0 and 104.0 for extrafamilial offenders. For both incest and extrafamilial offenders, the group with relatively high PCL-R scores and low IQ scores had the highest rate of sexual recidivism, as was the case for the sample as a whole. Kaplan-Meier survival analyses indicated that although the overall between-group differences were not significant among incest offenders, $\chi^2 (df = 3) = 6.10, ns$, the pairwise difference between the High PCL-R-High IQ and High PCL-R-Low IQ groups approached significance, $\chi^2 (df = 1) = 3.64, p < .06$. It is likely that the low base rate of recidivism among the incest offenders (4.8%) limited the potential for differences between the subgroups to reach statistical significance. The four groups of extrafamilial offenders differed

significantly in their rate of sexual recidivism, $\chi^2 (df = 3) = 13.40, p < .01$. Pairwise comparisons indicated that the High PCL-R-Low IQ group reoffended at a significantly faster rate than either the Low-PCL-R-High IQ group, $\chi^2 (df = 1) = 6.16, p < .05$, or the Low-PCL-R-Low IQ group, $\chi^2 (df = 1) = 7.41, p < .01$.

These results suggest that the tendency for higher sexual recidivism rates among offenders with higher PCL-R scores and lower intelligence occurred across the sample of child molesters, irrespective of type of victim relationship and despite differences in the intelligence, risk level, and recidivism rates between incest and extrafamilial offenders.

4.5 Discussion

The primary goal of this study was to examine the relationships between intelligence, psychopathy, and offending in a sample of child molesters who completed a prison-based treatment program. Consistent with previous research, regression analyses showed that psychopathy was related to every offending outcome: Sexual and non-sexual offense history, and sexual, violent, and general recidivism. Although intelligence was not significantly related to offending, the interaction between intelligence and psychopathy was significant for each recidivism variable. Comparisons between four groups obtained by performing median splits on psychopathy and intelligence clarified this interaction: The group with relatively high PCL-R and relatively low IQ scores was more than four times as likely to have been reconvicted of a sexual offense, and more than twice as likely to have been reconvicted of a violent or general offense, as any other group. Results were similar for extrafamilial and incest offenders when examined separately.

These findings show that intelligence may moderate the relationship between psychopathy, as measured by the PCL-R, and reoffending in child molesters, particularly for sexual recidivism: Those with relatively high PCL-R scores and high IQ received convictions for new sexual offenses at the same rate as those with low PCL-R scores. The present results thus extend Heilbrun's (1979) finding that IQ moderates the effects of psychopathy on violence to the prediction of child molester recidivism. In addition, no interaction effect was found between the Static-99 and intelligence on sexual or general recidivism. This suggests that the interaction with intelligence is specific to psychopathy as a personality construct, and does not apply to actuarial risk in general.

The interaction between intelligence and psychopathy was apparent in terms of the predictive validity of the PCL-R for recidivism, as measured by the area under the receiver operating characteristic curve (AUC). AUC values for the PCL-R were greater for offenders with relatively low IQ than for those with relatively high IQ for every type of recidivism (see Table 22). The AUC values for the PCL-R were always higher for low IQ offenders and the full sample than those for the Static-99, and also generally higher than those obtained in previous studies. For example, in a comparative validity study, Barbaree, Seto, Langton and Peacock (2001) reported an AUC value of .61 for the PCL-R predicting sexual recidivism. By contrast, in the current study the corresponding value was .73 for the full sample, increasing to .76 for relatively low IQ offenders. The good predictive validity of the PCL-R is especially noteworthy given the comparatively low distribution of scores in our data ($M = 8.2$, compared to $M = 16.1$ in Barbaree et al). The most salient difference between our sample and that studied by Barbaree et al. is that ours consisted entirely of child molesters, whereas Barbaree et al.'s also included rapists. Although it is unclear

whether these differences might have been responsible for the increased predictive validity of the PCL-R, our data suggest that the PCL-R can effectively predict recidivism in child molesters, especially if the offender's intelligence is taken into account.

An alternative way to view the interaction between psychopathy and intelligence is to note that recidivism rates were low for all offenders with relatively high IQ scores, even those in the high PCL-R group. According to this view, intelligence might be considered as a protective factor for recidivism, that is, that above-average intelligence can mitigate the risk associated with relatively high PCL-R scores. The implication is that measures of intelligence should be included in assessments of child molesters, and taken into account by clinicians particularly in cases where the PCL-R is used to predict risk. Our results suggest that although PCL-R scores are sensitive to risk for lower intelligence sex offenders, they have limited predictive validity for those of higher intelligence. By contrast, the lack of a significant interaction with the Static-99 suggests that this instrument is a valid predictor of sexual recidivism regardless of IQ. This may not extend to violent recidivism, for which an interaction between PCL-R and Static-99 scores was found. However as the Static-99 was specifically designed to assess risk for sexual recidivism, it would be unlikely to be a primary consideration in violence risk assessments. Further research would be useful to examine the effect of intelligence on predictions of violent reoffending, for example whether there might be an interaction between intelligence and a tool such as the Violence Risk Assessment Guide (VRAG, Harris, Rice, & Quinsey, 1993). This would be particularly important in light of the current findings, given that PCL-R scores are included as a component within the VRAG.

The present finding of low recidivism rates among relatively high IQ offenders also contributes to evidence that low intelligence is associated with pedophilia. For the relatively low IQ group, the mean IQ was 87.07 (the corresponding mean for the relatively high IQ group was 110.68). In a recent meta-analysis, Cantor, Blanchard, Robichaud and Christensen (2005) found that child molesters had lower IQ scores than other offenders and controls, with a mean of 95. Noting other findings of child molesters having greater frequencies of childhood brain injury and non-righthandedness (e.g., Blanchard et al., 2003; Cantor, Klassen, et al., 2005), Cantor, Klasson et al. proposed that the IQ deficit might be related to biological factors associated with the emergence of pedophilia, specifically a perturbation of brain development. Given the substantial evidence that psychopathy is associated with abnormal cognitive and autonomic functioning related to emotion processing (Herba et al., 2007), and recent reports that psychopathic personality in adulthood is correlated with psychophysiological reactivity at age three (Glenn, Raine, Venables, & Mednick, 2007), the present results encourage the speculation that a common biological antecedent might be responsible for the intelligence deficits and psychopathic personality features associated with pedophilia.

In summary, the interaction between psychopathy and intelligence found in this study supports past research (e.g., Heilbrun, 1979) suggesting that intelligence may moderate the effect of psychopathy on recidivism, extending this finding to sex offenders. The failure to find a corresponding interaction between intelligence and actuarial risk suggests that the effect is specific to psychopathy. These results have implications for the practice of risk assessment among child sex offenders, particularly the importance of taking and offender's level of intelligence into account.

General Discussion

The current volume of work represents the investigation of several pertinent issues within the general area of psychological treatment of sexual offenders. In this concluding section, an overview of the major results of the preceding empirical studies will be presented, followed by a discussion of the findings within the following key themes: risk assessment among sexual offenders; assessment of treatment outcome and its relationship with recidivism; and the influence of specific offender factors on assessments of treatment outcome and risk. Finally, areas for future investigation arising from the present findings will be proposed.

Overview of the Empirical Results of this Dissertation

Study 1 was an independent evaluation of the psychometric properties (in particular, the predictive validity) of the Violence Risk Scale: Sexual Offender Version (VRS:SO; Olver, Wong, Nicholaichuk, & Gordon, 2007), a recently-developed instrument designed to assess static and dynamic recidivism risk among sexual offenders. An initial validation study by the VRS:SO authors (Olver et al., 2007) found preliminary support for the inter-rater reliability, and concurrent validity of the measure, and factor analyses suggested a three-factor solution for the dynamic scale (sexual deviancy, criminality, and treatment responsiveness). Olver et al. (2007) also reported that the VRS:SO showed good predictive validity in terms of sexual recidivism (AUCs of .71 and .72 for pre-treatment and post-treatment total scores respectively), and that change across treatment on the dynamic scale was associated with decreased recidivism.

Study 1 represented an important contribution to ascertain whether the validity of the VRS:SO could be generalised to an independent sample by researchers not involved in the development of the measure. The results of Study 1 were positive in this respect: we found confirmatory support for the VRS:SO's inter-rater reliability, convergent validity, and predictive validity (AUCs of .77 and .79 respectively for pre-treatment and post-treatment total scores), thus extending the validity of the measure to a lower-risk child molester sample in a different cultural and geographical context. The VRS:SO dynamic scale was predictive of recidivism independently of static risk, and was also correlated with a four-factor psychometrically assessed "Deviance" framework previously developed by Allan, Grace, Rutherford, and Hudson (2007) on a larger sample of Kia Marama completers inclusive of the current sample. VRS:SO scores showed greater predictive validity than both static risk and the Deviance framework in the present dataset. Factor analyses of the VRS:SO dynamic scale in Study 1 suggested that the inclusion of a fourth factor (Self-Management) may provide a better description among groups resembling the current sample (i.e., lower-risk, child molesters), supported by theory and previous empirical findings (see Ward & Hudson, 1998, and Yates & Kingston, 2006). Overall, Study 1 provided further support for the inclusion of dynamic, changeable factors in sex offender risk assessments, in addition to evidence for the validity of the VRS:SO in particular. The relationship between changes on such factors as a result of treatment gain (as well as other conceptualisations of treatment outcome) and recidivism was investigated in Study 2.

In Study 2, three separate methods of assessing the proximal treatment outcome (in other words, the amount of change or gain achieved during treatment) of individual offenders were directly compared, in terms of advantages and

disadvantages of their use, and their external validity in terms of their association with recidivism. The first method involved change on a battery of relevant self-report psychometric tests administered to the sample prior to and following treatment. Problems with analysing raw change scores were highlighted, and two alternative options were pursued – analysing residual change after controlling for pre-treatment scores, and applying reliable and clinically significant change methodology – both of which also involved standardising the scores to allow for averages to be obtained for the four Deviance factors identified by Allan et al. (2007) and an overall average. The other two methods of assessing treatment outcome were based on therapists' impressions (contained in file information, primarily psychological treatment reports, and accessed retrospectively) – change on the VRS:SO; and Standard Goal Attainment Scaling scores (SGAS; Hogue, 1994), which indicate the extent to which the individual had attained important goals of treatment at the conclusion of the programme.

The results of Study 2 indicated that these three methods were correlated with each other, suggesting convergent reliability as measures of proximal treatment outcome. All three were also significantly predictive of reduced recidivism, such that those assessed as having benefited the most from treatment were less likely to have been reconvicted of a sexual offence than those assessed as having poorer treatment outcomes. These findings provided empirical support for the premise that dynamic risk variables are changeable (e.g., through treatment), and that change in the intended direction can be associated with reductions in recidivism. Previous findings on this issue have reportedly been limited to higher-risk offenders (Marques, Wiederanders, Day, Nelson, & van Ommeren, 2005; Olver et al., 2007), thus this finding from Study 2 is especially noteworthy given the lower risk profile of the sample. Clinically

significant change on the psychometric battery was able to predict incremental variance after controlling for the Static-99, while standardised residual psychometric change, VRS:SO change scores, and SGAS scores approached significance in this respect (certain sub-total factor scores did reach significance).

In terms of the comparative utility of these different methods, it was noted in Study 2 that the SGAS carries a major advantage of efficiency, and given the evidence for its association with recidivism can be recommended as a simple post-treatment screening tool of treatment outcome. Change on the VRS:SO showed comparable predictive validity, and there were several noted additional benefits of its use. These included the identification of specific criminogenic needs (i.e., treatment targets) salient to individual offenders, validated risk assessment protocols (Beggs & Grace, 2008 [Study 1, this volume]; Olver et al., 2007), and the simplicity of these functions all being contained within a single instrument. In contrast, the resource-intensive nature of psychometric batteries was noted (perhaps limiting their replicability to other services); nonetheless the predictive validity of change on the Kia Marama battery (in terms of reduced recidivism) was supported. Additionally (like the VRS:SO), psychometric testing can provide useful clinical information regarding treatment targets, and the Kia Marama battery has previously been shown to be linked to recidivism risk (Allan et al., 2007).

Studies 3 and 4 were investigations into the influence of particular offender characteristics on treatment outcome and recidivism. The potential for variables to interact with each other was also explored, specifically, via attempted replications of two previously reported interaction effects involving the personality construct of psychopathy (Heilbrun, 1979, 1982; Seto & Barbaree, 1999). In both studies, incest

and extrafamilial offenders were examined separately as well as together as a whole sample, in order to explore possible differences between the two groups.

Study 3 was an attempted replication of an interaction effect reported by Seto and Barbaree (1999), in which offenders who had relatively high PCL-R scores and were also rated as having positive outcomes of treatment were the most likely to commit a serious reoffence. Using SGAS scores (Hogue, 1994) as the measure of treatment outcome (previously validated in Study 2), and PCL-R scores (recorded as part of the standard assessment process at Kia Marama) as the measure of psychopathy, the results of Study 3 indicated a failure to replicate this interaction, consistent with findings reported by Barbaree (2005). Among offenders high on psychopathy, those with poor treatment outcome were more likely to reoffend sexually than those who scored low on psychopathy, while those who had more positive SGAS scores did not differ from low PCL-R scorers in terms of recidivism (it should be noted that the distribution of PCL-R scores in the present sample was relatively low, with a median score of 7 – “high” and “low” psychopathy were therefore used in these studies as relative terms based on a median split).

Several possible reasons for the discrepancy in results on this issue (e.g., Barbaree, 2005; Looman, Abracen, Serin, & Marquis, 2005; Seto & Barbaree, 1999) were discussed. These included the use of, in previous studies: the outcome measure of serious recidivism (a combined variable of sexual and/or violent recidivism) rather than sexual recidivism; measures of treatment outcome that pertain to treatment behaviour or compliance rather than purely measures of clinical gain (as posited by Looman et al., 2005); and potentially invalid measures of treatment outcome (i.e., those that are unrelated to recidivism, or appear to be associated with recidivism in an inverse direction as in Seto & Barbaree, 1999).

For the sub-sample of incest offenders in Study 3, positive treatment outcome was associated with increased risk of violent recidivism, with the highest rates for those who also scored relatively high on psychopathy (however this result was not obtained for sexual recidivism, or for extrafamilial offenders). Speculations regarding the cause of this differential finding included that incest offenders may have a greater likelihood of employing maladaptive coping strategies that are linked to violence, such as substance abuse; or that treatment of lower risk incest offenders is a counter-productive violation of the risk principle, with increased post-release violence a negative effect of this. In any case, the results of Study 3 suggested that the relationship between psychopathic traits, treatment outcome, and recidivism may depend on the offender sub-group as well as recidivism type; and that overall, sexual offenders should not be excluded from treatment on the basis of their PCL-R score, as it appeared that many were able to benefit from the programme and reduce their level of risk.

Study 4 involved the investigation of a potential interaction effect between psychopathy and intelligence on recidivism; an extension of previous research conducted by Heilbrun (1979, 1982). In this investigation, no differences were found between incest and extrafamilial offenders. A significant interaction effect was found among the whole sample – men with the combination of relatively high PCL-R scores and relatively low IQ scores were at a greatly increased risk of sexual reoffending, with a reconviction rate of more than four times the remainder of the sample. Offenders with relatively high PCL-R scores who had higher IQs in fact had the same rate of reconviction as those with lower PCL-R scores. These results extended those of Heilbrun (1979) regarding violent crime, and indicated that intelligence may moderate the relationship between psychopathy and sexual reoffending among child

molesters. Such an effect was not found between intelligence and Static-99 scores (Hanson & Thornton, 1999), suggesting that the interaction with intelligence is specific to psychopathy as a personality construct, and does not apply to the moderation of static risk in general. The implications of these findings as discussed in Study 4 included the possibility that the predictive validity of the PCL-R with regards to recidivism, previously supported in numerous studies including Barbaree, Seto, Langton, and Peacock (2001), may be dependent on the IQ of the offender. The predictive validity of the PCL-R among offenders with IQ scores higher than the current sample median of 99.5 was not supported in Study 4, and these findings suggested that intelligence should be assessed and taken into account in cases where the PCL-R is used to predict recidivism risk among child molesters.

Concluding Discussion – Key Themes

The findings from this dissertation as a whole have several general implications for the areas of assessing treatment outcome and risk among child sexual offenders. Several important original contributions to the field have been made, as well as support for previously reported empirical findings.

Regarding the assessment of recidivism risk among child sexual offenders, the importance of considering dynamic factors in addition to static risk is clear. The VRS:SO is an example of a risk instrument for sex offenders that incorporates assessment of both static and dynamic risk factors, and provides protocols for combining these to obtain an overall risk estimate. This measure has now been empirically validated as showing good predictive validity in two independent studies with adequate sample sizes and follow-up periods, in different cultural and

geographical contexts, and with different sample characteristics and risk profiles (the first in Canada by the VRS:SO authors, Olver et al., 2007; and the second in New Zealand – Study 1 of the current volume, Beggs & Grace, 2008). In the current dataset, this measure showed superior predictive validity with regard to sexual recidivism than both the Static-99 and a measure of Deviance based on scores on a battery of relevant psychometric tests. A great deal of important clinical information can be gained by assessing dynamic factors, and there is now much support for their independent and incremental predictive validity after controlling for static factors.

These findings are also consistent with Beech and Ward's model of "The Etiology of Risk" (Beech & Ward, 2004; Ward & Beech, 2004), in which static factors are considered to be merely historical markers for the psychological traits linked to vulnerability for sexual offending (i.e., stable dynamic risk factors). Risk assessments limited to the measurement of historical markers may therefore be valid, but would lack consideration of the specific psychological characteristics and clusters of these that may contribute to an individual's likelihood of sexual reoffending. Including such factors in a risk assessment would therefore be expected to increase the predictive accuracy (as has been shown). Beech and Ward's model also provides clarity as to the mechanisms by which sexual offending can occur. Specifically, acute risky states are hypothesised to be generated by the interaction of stable psychological vulnerabilities and contextual triggers. A thorough risk assessment based on an individualised case formulation using Beech and Ward's framework as a guide would not only be informative regarding appropriate treatment targets for an individual, but would also have important implications for risk management planning, due to the incorporation of triggering events or contextual risk factors.

For those offenders who complete a treatment programme, it is important to be able to assess the extent to which they have benefited such that post-treatment risk assessments can reflect this. Aside from this clinical application, there are numerous research applications of assessments of proximal treatment outcome, and it is important that measures for these purposes show external validity (i.e., are associated with the longer-term outcome of recidivism). Three separate approaches to assessing treatment outcome appear to be empirically associated with sexual recidivism: standardised residual change scores controlling for pre-treatment scores on a battery of relevant psychometric tests; change scores on the VRS:SO; and post-treatment ratings of the extent to which the goals of treatment have been attained (SGAS scores; Hogue, 1994). SGAS scores are simple and efficient to obtain at one point in time only (post-treatment), and yet showed comparable predictive validity to VRS:SO and psychometric change scores. Benefits of the more time-consuming methods include the useful pre-treatment clinical information gained regarding salient treatment targets for individual offenders, and in the case of the VRS:SO, not being reliant on offender self-reports and being contained within a single multi-purpose instrument. The changeable nature of dynamic factors are further supported, along with the association between changes on these factors across treatment and reduced sexual recidivism.

The presence of psychopathic traits is an example of a specific offender characteristic that can influence both response to treatment and risk of recidivism. It appears that psychopathy is associated with poorer treatment outcomes and increased recidivism, however it does not follow that such offenders should be excluded from treatment. On the contrary, some high scorers on the PCL-R (Hare, 1991) appear able to benefit from treatment, and participating in treatment does not appear to make such offenders worse in terms of sexual recidivism, particularly among men who have

offended against victims outside of their own family. Moreover, higher than average intelligence appears to be a protective factor against the increased likelihood of reoffending usually associated with psychopathic traits.

A limitation of the body of research presented here is that in all cases the sample consisted only of relatively low risk child molesters. It remains to be seen whether the findings could be generalised to samples of higher-risk child molesters or rapists. A cultural breakdown of the analyses described here would also be beneficial, particularly with regard to ascertaining the validity of the VRS:SO and SGAS on Maori offenders in particular as a special sub-group of the New Zealand offender population. More broadly, the present findings highlight the need for ongoing research into offender characteristics that can influence treatment response and recidivism risk, and the importance of taking such factors into account on an individualised basis in clinical settings, particularly with regard to risk assessments. Potential interactive effects between variables also require more exploration. Despite the many significant advancements made in recent years in this field, it would seem important for researchers to remain open minded about the very nature of sexual offender risk (e.g., see Doren, 2004), and continue to investigate ways to improve treatment approaches and reduce recidivism even further, for the benefit of society in general.

Ko te pae tawhiti whaia kia tata, ko te pae tata whakamana kia tina

Seek out distant horizons, and cherish those you attain

- Maori proverb of unknown origin kindly supplied by John Panirau,
Cultural Advisor and Kaumatua of the Kia Marama Special Treatment Unit

References

- Abel, G. G., Gore, D. K., Holland, C. L., Camp, N., Becker, J.V., & Rather, J. (1989). The measurement of the cognitive distortions in child molesters. *Annals of Sex Research, 2*, 135-153.
- Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse: A Journal of Research and Treatment, 19*, 347-367.
- Anderson, J., Martin, J., Mullen, P., Romans, S., & Herbison, P. (1993). Prevalence of childhood sexual abuse experiences in a community sample of women. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 911-919.
- Anderson, R. D., Gibeau, D., & D'Amora, D. A. (1995). The Sex Offender Treatment Rating Scale: Initial reliability data. *Sexual Abuse: A Journal of Research and Treatment, 7*, 221-227.
- Andrews, D. A., & Bonta, J. (2003). *The psychology of criminal conduct* (3rd ed.). Cincinnati: Anderson.
- Andrews, D.A., Bonta, J. & Hoge, R.D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.
- Bakker, L., Hudson, S. M., Wales, D. S., & Riley, D. (1998). *And there was light... Evaluating the Kia Marama treatment programme for New Zealand sex offenders against children*. New Zealand Department of Corrections Report. Available online at <http://www.corrections.govt.nz/public/pdf/research/kiamarama/kiamarama.pdf>

- Barbaree, H. E. (2005). Psychopathy, treatment behavior, and recidivism: An extended follow-up of Seto and Barbaree. *Journal of Interpersonal Violence, 20*, 1115-1131.
- Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28*, 490-521.
- Barrett, M., Wilson, R. J., & Long, C. (2003). Measuring motivation to change in sexual offenders from institutional intake to community treatment. *Sexual Abuse: A Journal of Research and Treatment, 15*, 269-283.
- Baumgartner, J. V., Scalora, M. J., & Huss, M. T. (2002). Assessment of the Wilson Sex Fantasy Questionnaire among child molesters and nonsexual forensic offenders. *Sexual Abuse: A Journal of Research & Treatment, 14*, 19-30.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*, 77-100.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.
- Beech, A. R. (1998). A psychometric typology of child abusers. *International Journal of Offender Therapy and Comparative Criminology, 42*, 319-339.
- Beech, A. R., Fisher, D. D., & Thornton, D. (2003). Risk assessment of sex offenders. *Professional Psychology: Research and Practice, 34*, 339-352.

- Beech, A., & Ford, H. (2006). The relationship between risk, deviance, treatment outcome and sexual reconviction in a sample of child sexual abusers completing residential treatment for their offending. *Psychology, Crime & Law, 12*, 685-701.
- Beech, A., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A Journal of Research and Treatment, 14*, 155-167.
- Beech, A. R., & Hamilton-Giachritsis. (2005). Relationship between therapeutic climate and treatment outcome in group-based sexual offender treatment programs. *Sexual Abuse: A Journal of Research and Treatment, 17*, 127-140.
- Beech, A. R., & Ward, T. (2004). The integration of etiology and risk in sexual offenders: A theoretical framework. *Aggression and Violent Behaviour, 10*, 31-63.
- Beggs, S. M., & Grace, R. C. (2008). *Assessment of dynamic risk factors: An independent validation study of the Violence Risk Scale: Sexual Offender Version*. Manuscript submitted for publication (Study 1, this volume).
- Bickley, J. A., & Beech, A. R. (2003). Implications for treatment of sexual offenders of the Ward and Hudson model of relapse. *Sexual Abuse: A Journal of Research and Treatment, 15*, 121-134.
- Blanchard, R., Kuban, M. E., Klassen, P., Dickey, R., Christensen, B. K., Cantor, J. M., et al. (2003). Self-reported head injuries before and after age 13 in pedophilic and nonpedophilic men referred for clinical assessment. *Archives of Sexual Behavior, 32*, 573-581.

- Boles, S. M., & Miotto, K. (2003). Substance and violence: A review of the literature. *Aggression and Violent Behavior, 8*, 155-174.
- Brown, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 445-455). Newbury Park, CA: Sage.
- Browne, K. D., Foreman, L., & Middleton, D. (1998). Predicting treatment drop-out in sex offenders. *Child Abuse Review, 7*, 402-419.
- Burk, L. R., & Burkhart, B. R. (2003). Disorganized attachment as a diathesis for sexual deviance: Developmental experience and the motivation for sexual offending. *Aggression and Violent Behavior, 8*, 487-511.
- Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology, 38*, 217-230.
- Cantor, J. M., Blanchard, R., Robichaud, L. K., & Christensen, B. K. (2005). Quantitative reanalysis of aggregate data on IQ in sexual offenders. *Psychological Bulletin, 131*, 555-568.
- Cantor, J. M., Klassen, P. E., Dickey, R., Christensen, B. K., Kuban, M. E., Blak, T., et al. (2005). Handedness in pedophilia and hebephilia. *Archives of Sexual Behavior, 34*, 447-459.
- Check, J. V. P. (1985). *The hostility towards women scale*. Unpublished doctoral dissertation, University of Manitoba, Winnipeg.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Craig, L. A., Browne, K. D., & Stringer, I. (2003a). Risk scales and factors predictive of sexual offence recidivism. *Trauma Violence & Abuse, 4*, 45-69.

- Craig, L. A., Browne, K. D., & Stringer, I. (2003b). Treatment and sexual offence recidivism. *Trauma Violence & Abuse, 4*, 70-89.
- Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2004). Limitations in actuarial risk assessment of sexual offenders: A methodological note. *British Journal of Forensic Practice, 6*, 16-32.
- Craissati, J., & Beech, A. R. (2003). A review of dynamic variables and their relationship to risk prediction in sex offenders. *Journal of Sexual Aggression, 9*, 41-55.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change" – or should we? *Psychological Bulletin, 74*, 68-80.
- Dean, K. E., & Malamuth, N. M. (1997). Characteristics of men who aggress sexually and of men who imagine aggressing: Risk and moderating variables. *Journal of Personality and Social Psychology, 72*, 449-455.
- Descutner, C. J., & Thelen, M. H. (1991). Development and validation of a Fear-of-Intimacy Scale. *Psychological Assessment, 3*, 218-225.
- Dhawan, S., & Marshall, W. L. (1996). Sexual abuse histories of sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 8*, 7-15.
- Doi, S., & Thelen, M. H. (1993). The Fear-of-Intimacy Scale: Replication and extension. *Psychological Assessment, 5*, 377-383.
- Donato, R., & Shanahan, M. (2001). The economics of child sex-offender rehabilitation programs: Beyond Prentky & Burgess. *American Journal of Orthopsychiatry, 71*, 131-139.
- Doren, D. M. (2004). Toward a multidimensional model for sexual recidivism risk. *Journal of Interpersonal Violence, 19*, 835-856.

- Doren, D. M. (2006). Recidivism risk assessments: Making sense of controversies. In W. L. Marshall, Y. M. Fernandez, L. E. Marshall & G. A. Serran (Eds.), *Sexual offender treatment: Controversial issues* (pp. 3-15). Chinchester: Wiley.
- Dowden, C., Antonowicz, D., & Andrews, D. A. (2003). The effectiveness of relapse prevention with offenders: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology*, 47, 516-528.
- Eccles, A., & Marshall, W. L., (1999). Relapse prevention. In W. L. Marshall, D. Anderson, & Y. Fernandez, *Cognitive Behavioural Treatment of Sexual Offenders* (pp. 93-109). Chinchester: Wiley.
- Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health Notebook*, 1, 70-72.
- Fergusson, D. M., & Horwood, L. J. (1995). Early disruptive behavior, IQ, and later school achievement and delinquent behavior. *Journal of Abnormal Child Psychology*, 23, 183-199.
- Fernandez, Y. M., Shingler, J., & Marshall, W. L. (2006). Putting "behavior" back into the cognitive-behavioral treatment of sexual offenders. In W. L. Marshall, Y. M. Fernandez, L. E. Marshall & G. A. Serran (Eds.), *Sexual offender treatment: Controversial issues* (pp. 211-224). Chinchester: Wiley.
- Finkelhor, D. (1994). The international epidemiology of child sexual abuse. *Child Abuse & Neglect*, 18, 409-417.
- Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Larose, M. R., & Curry, S. (1999). Prediction of recidivism in incest offenders. *Journal of Interpersonal Violence*, 14, 511-531.

- Fisher, D., Beech, A., & Browne, K. (1999). Comparison of sex offenders to non-sex offenders on selected psychological measures. *International Journal of Offender Therapy and Comparative Criminology, 43*, 473-491.
- Gambrill, E. D., & Richey, C. A. (1975). An assertion inventory for use in assessment and research. *Behavior Therapy, 6*, 550-561.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*, 575-608.
- Ginsburg, J. I. D., Mann, R. E., Rotgers, F., & Weekes, J. R. (2002). Motivational interviewing with criminal justice populations. In W. R. Miller & S. Rollnick (Eds.), *Motivational interviewing: Preparing people for change* (2nd ed., pp. 333-346). New York: Guilford Press.
- Glenn, A. L., Raine, A., Venables, P. H., & Mednick, S. A. (2007). Early temperamental and psychophysiological precursors of adult psychopathic personality. *Journal of Abnormal Psychology, 116*, 508-518.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*, 293-323.
- Guay, J.-P., Ouimet, M., & Proulx, J. (2005). On intelligence and crime: A comparison of incarcerated sex offenders and serious non-sexual violent criminals. *International Journal of Law and Psychiatry, 28*, 405-417.
- Guy, L. S., Edens, J. F., Anthony, C., & Douglas, K. S. (2005). Does psychopathy predict institutional misconduct among adults? A meta-analytic investigation. *Journal of Consulting & Clinical Psychology, 73*, 1056-1064.

- Hall, G. C. N. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology, 63*, 802-809.
- Hanson, R. K. (1997a). *The development of a brief actuarial risk scale for sexual offense recidivism (User Report 1997-04)*. Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K. (1997b). How to know what works with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 9*, 129-145.
- Hanson, R. K. (1998). What do we know about sex offender risk assessment? *Psychology, Public Policy, and Law, 4*, 50-72.
- Hanson, R. K. (2000). What is so special about relapse prevention? In D. R. Laws, S. M. Hudson & T. Ward (Eds.), *Remaking relapse prevention with sex offenders: A sourcebook* (pp. 27-38). Thousand Oaks, CA: Sage.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.
- Hanson, R. K., Cox, B. J., & Woszczyzna, C. (1991). Assessing treatment outcome for sexual offenders. *Annals of Sex Research, 4*, 177-208.
- Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., et al. (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*, 169-194.
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene?: Dynamic predictors of sexual assault recidivism. *Criminal Justice and Behavior, 27*, 6-35.

- Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *13*, 105-122.
- Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (No. 2007-05): Public Safety Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology*, *73*, 1154-1163.
- Hanson, R. K., Steffy, R. A., & Gauthier, R. (1993). Long-term recidivism of child molesters. *Journal of Consulting and Clinical Psychology*, *61*, 646-652.
- Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders (User report 1999-02)*: Ottawa: Department of the Solicitor General of Canada.
- Hare, R. D. (1991). *Manual for the Revised Psychopathy Checklist*. Toronto: Multi-Health Systems.
- Hare, R.D. (2003). *The Hare Psychopathy Checklist – Revised* (2nd ed.). Toronto: Canada. Multi-Health Systems.
- Hare, R. D., Clark, D., Grann, M., & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences & the Law. Special Issue: International Perspectives on Psychopathic Disorders*, *18*, 623-645.
- Hare, R. D., Hart, S. D., & Harpur, T. J. (1991). Psychopathy and the DSM-IV criteria for antisocial personality disorder. *Journal of Abnormal Psychology*, *100*, 391-398.

- Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *STATIC-99 coding rules revised – 2003*. Ottawa: Department of the Solicitor General of Canada.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315-335.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-425.
- Hartley, C. C. (2001). Incest offenders' perceptions of their motives to sexually offend within their past and current life context. *Journal of Interpersonal Violence, 16*, 459-475.
- Heilbrun, A. B. (1979). Psychopathy and violent crime. *Journal of Consulting and Clinical Psychology, 47*, 509-516.
- Heilbrun, A. B. (1982). Cognitive models of criminal violence based upon intelligence and psychopathy levels. *Journal of Consulting and Clinical Psychology, 50*, 546-557.
- Hemphill, J. F., Hare, R. D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology, 3*, 139-170.
- Herba, C. M., Hodgins, S., Blackwood, N., Kumari, V., Naudts, K. H., & Phillips, M. (2007). The neurobiology of psychopathy: A focus on emotion processing. In H. Herve & J. C. Yuille (Eds.), *The psychopath: Theory, research, and practice* (pp. 253-283). Mahwah, NJ: Lawrence Erlbaum.
- Hirschi, T., & Hindelang, M. J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review, 42*, 571-587.

- Hogue, T. E. (1994). Goal attainment scaling: A measure of clinical impact and risk assessment. *Issues in Criminological & Legal Psychology, 21*, 96-102.
- Holland, T. R., Beckett, G. E., & Levi, M. (1981). Intelligence, personality, and criminal violence: A multivariate analysis. *Journal of Consulting and Clinical Psychology, 49*, 106-111.
- Hudson, S. M., Marshall, W. L., Ward, T., Johnston, P. W., & Jones, R. L. (1995). Kia Marama: A cognitive-behavioural program for incarcerated child molesters. *Behaviour Change, 12*, 69-80.
- Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: The Kia Marama evaluation. *Sexual Abuse: A Journal of Research and Treatment, 14*, 101-117.
- Hudson, S. M., Wales, D. S., & Ward, T. (1998). Kia Marama: A treatment program for child molesters in New Zealand. In W. L. Marshall, Y. M. Fernandez, S. M. Hudson & T. E. Ward (Eds.), *Sourcebook of treatment programs for sexual offenders* (pp. 17-28). New York: Plenum Press.
- Hudson, S. M., & Ward, T. (1996). Relapse prevention: Future directions. *Sexual Abuse: A Journal of Research and Treatment, 8*, 249-256.
- Hudson, S. M., Ward, T., & McCormack, J. C. (1999). Offense pathways in sexual offenders. *Journal of Interpersonal Violence, 14*, 779-798.
- Kendall-Tackett, K. A., Williams, L. M., & Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin, 113*, 164-180.
- Kingston, D. A. (2003). *Violence Risk Scale: Sex Offender Version: A re-validation of a scale to assess re-offence risk among sexual offenders*. Unpublished honours project, Carleton University, Ottawa.

- Krause, M. S., Howard, K. I., & Lutz, W. (1998). Exploring individual change. *Journal of Consulting and Clinical Psychology, 66*, 838-845.
- Langton, C. M., Barbaree, H. E., Harkins, L., & Peacock, E. J. (2006). Sex offenders' response to treatment and its association with recidivism as a function of psychopathy. *Sexual Abuse: A Journal of Research and Treatment, 18*, 99-120.
- Laws, D.R. (Ed.). (1989). *Relapse prevention with sex offenders*. New York, NY: Guilford Press.
- Laws, D. R. (2003). The rise and fall of relapse prevention. *Australian Psychologist, 38*, 22-30.
- Laws, D. R., Hudson, S. M., & Ward, T. (Eds.). (2000). *Remaking relapse prevention with sex offenders*. Thousand Oaks, CA: Sage.
- Lawson, J. S., Marshall, W. L., & McGrath, P. (1979). The Social Self-Esteem Inventory. *Educational and Psychological Measurement, 39*, 803-811.
- Levenson, J. S., & Macgowan, M. J. (2004). Engagement, denial, and treatment progress among sex offenders in group therapy. *Sexual Abuse: A Journal of Research and Treatment, 16*, 49-63.
- Leventhal, J. M. (1998). Epidemiology of sexual abuse of children: Old problems, new directions. *Child Abuse and Neglect, 22*, 481-491.
- Looman, J., Abracen, J., Maillet, G., & DiFazio, R. (1998). Phallometric nonresponding in sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 10*, 325-336.
- Looman, J., Abracen, J., Serin, R., & Marquis, P. (2005). Psychopathy, treatment change, and recidivism in high-risk, high-need sexual offenders. *Journal of Interpersonal Violence, 20*, 549-568.

- Looman, J., Dickie, I., & Abracen, J. (2005). Responsivity issues in the treatment of sexual offenders. *Trauma, Violence, & Abuse, 6*, 330-353.
- Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1*, 117-146.
- Mahoney, M. J. (1991). *Human change processes: The scientific foundations of psychotherapy*. New York: Basic Books.
- Malamuth, N. M., Sockloskie, R. J., Koss, M. P., & Tanaka, J. S. (1991). Characteristics of aggressors against women: Testing a model using a national sample of college students. *Journal of Consulting and Clinical Psychology, 59*, 670-681.
- Mann, R. E. (2000). Managing resistance and rebellion in relapse prevention. In D. R. Laws, S. M. Hudson & T. Ward (Eds.), *Remaking relapse prevention for sex offenders* (pp. 187-200). Thousand Oaks, CA: Sage.
- Mann, R. E., Webster, S. D., Schofield, C., & Marshall, W. L. (2004). Approach versus avoidance goals in relapse prevention with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 16*, 65-75.
- Marques, J. K. (1999). How to answer the question "Does sex offender treatment work?" *Journal of Interpersonal Violence, 14*, 437-451.
- Marques, J. K., Day, D. M., Nelson, C., & West, M. A. (1994). Effects of cognitive-behavioral treatment on sex offender recidivism: Preliminary results of a longitudinal study. *Criminal Justice and Behavior, 21*, 28-54.
- Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final

- results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A Journal of Research and Treatment*, 17, 79-107.
- Marshall, W. L. (2005). Therapist style in sexual offender treatment: Influence on indices of change. *Sexual Abuse: A Journal of Research and Treatment*, 17, 109-116.
- Marshall, W. L., & Anderson, D. (1996). An evaluation of the benefits of relapse prevention programs with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 8, 209-221.
- Marshall, W. L., Anderson, D., & Fernandez, Y. (1999). *Cognitive behavioural treatment of sexual offenders*. Chichester: Wiley.
- Marshall, W. L., Hudson, S. M., Jones, R., & Fernandez, Y. M. (1995). Empathy in sex offenders. *Clinical Psychology Review*, 15, 99-113.
- Marshall, W. L., & Marshall, L. E. (2000). The origins of sexual offending. *Trauma, Violence, and Abuse*, 1, 250-263.
- Marshall, W. L., Marshall, L. E., Serran, G. A., & Fernandez, Y. M. (2006). *Treating sexual offenders: An integrated approach*. New York, NY: Routledge.
- Marshall, W. L., & McGuire, J. (2003). Effect sizes in the treatment of sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, 47, 653-663.
- Marshall, W. L., & Serran, G. A. (2004). The role of the therapist in offender treatment. *Psychology, Crime & Law*, 10, 309-320.
- Marshall, W. L., Serran, G. A., Fernandez, Y. M., Mulloy, R., Mann, R. E., & Thornton, D. (2003). Therapist characteristics in the treatment of sexual offenders: Tentative data on their relationship with indices of behaviour change. *Journal of Sexual Aggression*, 9, 25-30.

- Marshall, W. L., Serran, G., Moulden, H., Mulloy, R., Fernandez, Y. M., Mann, R. E., et al. (2002). Therapist features in sexual offender treatment: Their reliable identification and influence on behaviour change. *Clinical Psychology & Psychotherapy, 9*, 395-405.
- Miethe, T. D., Olson, J., & Mitchell, O. (2006). Specialization and persistence in the arrest histories of sex offenders: A comparative analysis of alternative measures and offense types. *Journal of Research in Crime and Delinquency, 43*, 204-229.
- Morrison, Z., Quadara, A., & Boyd, C. (2007). "Ripple effects" of sexual assault. *Australian Centre for the Study of Sexual Assault Issues, 7*, 1-32.
- Mulloy, R., & Marshall, W. L. (1999). Social functioning. In W. L. Marshall, D. Anderson, & Y. Fernandez, *Cognitive Behavioural Treatment of Sexual Offenders* (pp. 93-109). Chichester: Wiley.
- Nowicki, S., & Duke, M. P. (1974). A locus of control scale for noncollege as well as college adults. *Journal of Personality Assessment, 38*, 136-137.
- Nowicki, S., & Duke, M. P. (1983). The Nowicki-Strickland Life-Span Locus of Control Scales: Construct validation. In H. M. Lefcourt (Ed.), *Research with the locus of control construct: Vol. 2. Developments and social problems* (pp. 9-49). New York: Academic Press.
- Ogloff, J. R. P., Wong, S., & Greenwood, A. (1990). Treating criminal psychopaths in a Therapeutic Community program. *Behavioral Sciences and the Law, 8*, 181-190.
- Olver, M. E. (2003). *The development and validation of the Violence Risk Scale: Sexual Offender Version (VRS:SO) and its relationship to psychopathy and*

treatment attrition. Unpublished doctoral dissertation, University of Saskatchewan, Saskatoon.

- Olver, M. E., Wong, S. C. P., Nicholaichuk, T. P., & Gordon, A. E. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender Version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment, 19*, 318-329.
- Plaud, J., & Bigwood, S. (1997). A multivariate analysis of the sexual fantasy themes of college men. *Journal of Sex and Marital Therapy, 23*, 221-230.
- Porter, S., Fairweather, D., Drugge, J., Hugues, H., Birt, A., & Boer, D. P. (2000). Profiles of psychopathy in incarcerated sexual offenders. *Criminal Justice & Behavior, 27*, 216-233.
- Prentky, R. A., Lee, A. F. S., Knight, R. A., & Cerce, D. (1997). Recidivism rates among child molesters and rapists: A methodological analysis. *Law and Human Behavior, 21*, 635-659.
- Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist, 47*, 1102-1114.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington D.C.: American Psychological Association.
- Quinsey, V. L., Khanna, A., & Malcolm, P. B. (1998). A retrospective evaluation of the Regional Treatment Centre Sex Offender Treatment Program. *Journal of Interpersonal Violence, 13*, 621-644.
- Reynolds, C. R., Willson, V. L., & Clark, P. L. (1983). A four-test short form of the WAIS-R for clinical screening. *Clinical Neuropsychology, 5*, 111-116.

- Rice, M. E., & Harris, G. T. (2003). The size and sign of treatment effects in sex offender therapy. *Annals of the New York Academy of Sciences*, 989, 428-440.
- Rice, M. E., Harris, G. T., & Cormier, C. A. (1992). An evaluation of a maximum security therapeutic community for psychopaths and other mentally disordered offenders. *Law & Human Behavior*, 16, 399-412.
- Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA Loneliness Scale. *Journal of Personality and Social Psychology*, 39, 472-480.
- Rutherford, B., & Grace, R. (2004, August). *Treatment outcome for sex offenders at Kia Marama: Benefits of bicultural practice?* Paper presented at the meeting of the New Zealand Psychological Society, Wellington.
- Scalora, M. J., & Garbin, C. (2003). A multivariate analysis of sex offender recidivism. *International Journal of Offender Therapy and Comparative Criminology*, 47, 309-323.
- Seager, J.A., Jellicoe, D. & Dhaliwal, G.K. (2004). Refusers, dropouts, and completers: Measuring sex offender treatment efficacy. *International Journal of Offender Therapy and Comparative Criminology*, 48, 600-612.
- Seggar, L. B., Lambert, M. J., & Hansen, N. B. (2002). Assessing clinical significance: Application to the Beck Depression Inventory. *Behavior Therapy*, 33, 253-269.
- Serran, G., Fernandez, Y., Marshall, W. L., & Mann, R. E. (2003). Process issues in treatment: Application to sexual offender programs. *Professional Psychology: Research & Practice*, 34, 368-375.
- Seto, M. C., & Barbaree, H. E. (1999). Psychopathy, treatment behavior, and sex offender recidivism. *Journal of Interpersonal Violence*, 14, 1235-1248.

- Shaw, J. A. (2004). The legacy of child sexual abuse. *Psychiatry: Interpersonal and Biological Processes, 67*, 217-221.
- Spencer, A. (1998). Peterhead Prison program. In W. L. Marshall, Y. M. Fernandez, S. M. Hudson & T. Ward (Eds.), *Sourcebook of treatment programs for sexual offenders* (pp. 29-46). New York, NY: Plenum Press.
- Spielberger, C. D. (1983). *State-Trait Anxiety Inventory*. Palo Alto, CA.: Consulting Psychologists Press.
- Spielberger, C. D. (1988). *State-Trait Anger Expression Inventory (STAXI) Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Stirpe, T. S., Wilson, R. J., & Long, C. (2001). Goal attainment scaling with sexual offenders: A measure of clinical impact at posttreatment and at community follow-up. *Sexual Abuse: A Journal of Research and Treatment, 13*, 65-77.
- Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment, 14*, 139-153.
- Tierney, D. W., & McCabe, M. P. (2001). An evaluation of self-report measures of cognitive distortions and empathy among Australian sex offenders. *Archives of Sexual Behavior, 30*, 495-519.
- Walsh, Z., Swogger, M. T., & Kosson, D. S. (2004). Psychopathy, IQ, and violence in European American and African American county jail inmates. *Journal of Consulting & Clinical Psychology, 72*, 1165-1169.
- Ward, T. & Beech, A. R. (2004). The etiology of risk: A preliminary model. *Sexual Abuse: A Journal of Research and Treatment, 16*, 271-284.
- Ward, T., & Gannon, T. A. (2006). Rehabilitation, etiology, and self-regulation: The comprehensive good lives model of treatment for sexual offenders. *Aggression and Violent Behavior, 11*, 77-94.

- Ward, T., & Hudson, S. M. (1998). A model of the relapse process in sexual offenders. *Journal of Interpersonal Violence, 13*, 700-725.
- Ward, T., Polaschek, D. L. L., & Beech, A. R. (2006). *Theories of sexual offending*. Chichester: Wiley.
- Ward, T. & Stewart, C. A. (2003). The treatment of sex offenders: Risk management and good lives. *Professional Psychology: Research and Practice, 34*, 353-360.
- Ward, T., Vess, J., Collie, R. M., & Gannon, T. A. (2006). Risk management or goods promotion: The relationship between approach and avoidance goals in treatment for sex offenders. *Aggression & Violent Behavior, 11*, 378-393.
- Wechsler, D. (1999). *The Wechsler Abbreviated Scale of Intelligence*. London: The Psychological Corporation.
- Wilson, G. (1978). *The secrets of sexual fantasy*. London: Dent.
- Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law, 12*, 279-309.
- Wong, S., Olver, M., Nicholaichuk, T., & Gordon, A. (2006). *Violence Risk Scale: Sexual Offender Version* (Manual provided by authors).
- Yates, P. M., & Kingston, D. A. (2006). The self-regulation model of sexual offending: The relationship between offence pathways and static and dynamic sexual offence risk. *Sexual Abuse: A Journal of Research and Treatment, 18*, 259-270.